GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 11:46:04 ; Search time 1300.69 Seconds
(without alignments)
253.213 Million cell updates/sec

Title: US-09-101-423B-8
Perfect score: 47
Sequence: 1 agctataatgcggccgcata.....tgatcggccgcaagcttgga 47

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*
44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estpl6:*
118: em_estpl7:*
119: em_estpl8:*
120: em_estro14:*
121: em_estro15:*
122: em_estro16:*
123: em_estro17:*
124: em_estro18:*
125: em_estro19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est75:*
136: gb_est76:*
137: gb_est77:*
138: gb_est78:*
139: gb_est79:*
140: gb_est80:*
141: gb_est81:*
142: gb_est82:*
143: gb_est83:*
144: gb_est84:*
145: gb_est85:*
146: gb_est86:*
147: gb_est87:*
148: gb_gss1:*
149: gb_gss2:*
150: gb_gss3:*
151: gb_gss4:*
152: em_gss1:*
153: em_gss2:*
154: em_gss3:*
155: em_gss4:*
156: gb_gss5:*
157: gb_gss6:*
158: gb_gss7:*
159: gb_gss8:*
160: gb_gss9:*
161: gb_gss10:*
162: gb_gss11:*
163: gb_gss12:*
164: gb_gss13:*
165: gb_gss14:*
166: gb_gss15:*
167: gb_gss16:*
168: gb_gss17:*
169: gb_gss18:*
170: gb_gss19:*
171: gb_gss20:*
172: gb_gss21:*
173: gb_gss22:*
174: gb_gss23:*
175: gb_gss24:*
176: em_gss5:*
177: em_gss6:*
178: em_gss7:*
179: em_gss8:*
180: em_gss9:*
181: em_gss10:*
182: em_gss11:*
183: em_gss12:*
184: em_gss13:*
185: em_gss14:*
186: em_gss15:*
187: em_gss16:*
188: em_gss17:*
189: em_gss18:*
190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	23.4	49.8	313	18	AI283439	AI283439 qh93e10.x
	2	23.4	49.8	520	4	AA236621	AA236621 zs43g09.r
c	3	23.2	49.4	558	163	AQ841408	AQ841408 T136718b
c	4	23.2	49.4	1054	134	BE036773	BE036773 MP05D02 M
c	5	22.4	47.7	705	97	AW977483	AW977483 EST389592
c	6	22.2	47.2	528	134	BE062593	BE062593 QV1-BT026
c	7	22.2	47.2	625	21	AI514427	AI514427 LD40932.5
	8	22	46.8	267	14	AA952587	AA952587 TENS1771
	9	22	46.8	313	23	AI667900	AI667900 TENGO871
c	10	22	46.8	447	1	AA063113	AA063113 zf68e10.s
c	11	22	46.8	475	17	AI187524	AI187524 EST316 Ma
	12	22	46.8	892	109	BE535565	BE535565 601059673
	13	21.8	46.4	436	37	AV624476	AV624476 AV624476
c	14	21.8	46.4	558	170	AZ342872	AZ342872 1M0076E01
c	15	21.8	46.4	610	87	AW181098	AW181098 MgA0062r
c	16	21.8	46.4	640	87	AW181091	AW181091 MgA0032r
	17	21.6	46.0	541	159	AQ578848	AQ578848 nbxb0093F
c	18	21.6	46.0	614	169	AZ240289	AZ240289 RPCI-23-3
	19	21.6	46.0	650	7	AA438826	AA438826 LD13184.5
	20	21.6	46.0	677	138	BE976052	BE976052 bs46f11.x
	21	21.6	46.0	707	97	AW940801	AW940801 GH20963.3
c	22	21.4	45.5	288	40	AW144909	AW144909 707013F10
c	23	21.4	45.5	412	89	AW331456	AW331456 707013F10
c	24	21.4	45.5	482	40	AW147159	AW147159 707013F10
c	25	21.4	45.5	947	193	CNS05GOO	AL336561 Tetraodon
	26	21.2	45.1	263	36	AV425024	AV425024 AV425024
	27	21.2	45.1	266	36	AV426457	AV426457 AV426457
	28	21.2	45.1	331	89	AW330028	AW330028 TENU4569
	29	21.2	45.1	403	36	AV409533	AV409533 AV409533
	30	21.2	45.1	463	36	AV422473	AV422473 AV422473
	31	21.2	45.1	554	3	AA196000	AA196000 zp99f07.s
	32	21.2	45.1	571	160	AQ615803	AQ615803 HS_5145_B
	33	21.2	45.1	813	110	BE640770	BE640770 Cri2_1_A1
	34	21.2	45.1	912	97	AW925378	AW925378 HVSMEg000
	35	21	44.7	220	91	AW491261	AW491261 UI-M-BH3-
	36	21	44.7	244	137	BE943769	BE943769 UI-M-BH3-
	37	21	44.7	272	91	AW491371	AW491371 UI-M-BH3-
c	38	21	44.7	351	107	BE363902	BE363902 PI1_10_G0
c	39	21	44.7	352	24	AI768069	AI768069 wi46d08.x
c	40	21	44.7	366	143	N77089	N77089 yv51c08.r1
c	41	21	44.7	372	14	AB009130	AB009130 AB009130
c	42	21	44.7	391	36	AV420471	AV420471 AV420471
c	43	21	44.7	433	40	AW144910	AW144910 707013F10
c	44	21	44.7	448	110	BE647694	BE647694 UI-M-BH1-
c	45	21	44.7	463	39	AW077673	AW077673 fj36b05.y

ALIGNMENTS

RESULT 1
AI283439
LOCUS AI283439 313 bp mRNA EST 23-NOV-1998
DEFINITION qh93e10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1854570 3' similar to TR:P76904 P76904 SIMILAR TO ;, mRNA sequence.
ACCESSION AI283439
VERSION AI283439.1 GI:3921672
KEYWORDS EST.

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 297.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1854570"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT       67 a     96 c    104 g     46 t
ORIGIN
```

```
  Query Match             49.8%;  Score 23.4;  DB 18;  Length 313;
  Best Local Similarity   73.2%;  Pred. No. 31;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       6 taatgcggccgcatattcggcctgatcggccgcaagcttgg 46
           || ||||||| |  |||||||| | ||| | |  |||| ||
Db      35 TACTGCGGCCACCAATTCGGCCAGTTCGCCGGGCAGCTGGG 75
```

```
RESULT   2
AA236621
LOCUS       AA236621      520 bp    mRNA            EST       07-AUG-1997
DEFINITION  zs43g09.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:688000 5'
            similar to TR:G1079677 G1079677 LPE5P. ;, mRNA sequence.
ACCESSION   AA236621
VERSION     AA236621.1  GI:1860641
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 618   Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Homo sapiens"
                     /db_xref="GDB:5592156"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:688000"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT      109 a    170 c    149 g     92 t
ORIGIN
```

```
  Query Match             49.8%;  Score 23.4;  DB 4;  Length 520;
  Best Local Similarity   73.2%;  Pred. No. 33;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       7 aatgcggccgcatattcggcctgatcggccgcaagcttgga 47
             ||| | ||||||  ||| |||||  | |||||  || ||||
Db      22 AATTCTGCCGCAAGTTCCGCCTGGACTGCCGCTGGCATGGA 62
```

```
RESULT   3
AQ841408/c
LOCUS       AQ841408      558 bp    DNA             GSS       27-SEP-1999
DEFINITION  T136718b shotgun sub-library of BAC clone 16K14 Medicago truncatula
            genomic clone 16-K-14-022, DNA sequence.
ACCESSION   AQ841408
VERSION     AQ841408.1  GI:5930433
KEYWORDS    GSS.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Kim,D., Choi,H., Peng,H., Ellis,L. and Cook,DR.
  TITLE     BAC survey sequencing of Medicago truncatula (1999c)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Cook DR
            The Crop Biotechnology Center
            Texas A&M University
            Department of Plant Pathology and Microbiology, Rm 120 L.F.
            Peterson Bldg, College Station, TX 77843-2132, USA
            Tel: 409 845 8743
            Fax: 409 862 4790
            Email: dcook@ppserver.tamu.edu
            Other name: BSC-19-22; date: 6/8/99; Submitted to the Database of
            Genome Survey Sequences (GSS) on 09/27/99; More information is
            available at 'http://chrysie.tamu.edu/medicago'.
            Seq primer: pUC-C
            Class: BAC subclone.
FEATURES             Location/Qualifiers
```

```
    source          1. .558
                    /organism="Medicago truncatula"
                    /cultivar="genotype A17"
                    /db_xref="taxon:3880"
                    /clone="16-K-14-022"
                    /clone_lib="shotgun sub-library of BAC clone 16K14"
                    /note="Vector: pUC18; BAC survey sequences were obtained
                    from sheared BAC DNA subcloned into the SmaI site of
                    pUC18. The template DNA for sequencing was obtained by PCR
                    using universal primers. Sequencing reactions were primed
                    from the PUC-C primer site (CAGGAAACAGCTATGACCATGATTACGA)
                    in the pUC18 polylinker."
BASE COUNT      168 a     76 c     95 g    219 t
ORIGIN


  Query Match             49.4%;  Score 23.2;  DB 163;  Length 558;
  Best Local Similarity   70.5%;  Pred. No. 39;
  Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      4 tataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
          ||||||||  || | |||   |||||  | ||| ||||||| | |
Db    487 TATAATGCTTCCCCTTATAAAGCCTGTGCAGCCACAAGCTAGCA 444


RESULT   4
BE036773/c
LOCUS       BE036773      1054 bp    mRNA            EST       07-JUN-2000
DEFINITION  MP05D02 MP Mesembryanthemum crystallinum cDNA 5', mRNA sequence.
ACCESSION   BE036773
VERSION     BE036773.1  GI:8331789
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
            Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 1054)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu
            Best blastx match: 'emb|CAB38794.1| (AL035678) putative protein
            [Arabidopsis thaliana] 267 9e-71'.
            Insert Length: 1   Std Error: 0.00.
FEATURES              Location/Qualifiers
     source           1. .1054
                      /organism="Mesembryanthemum crystallinum"
                      /db_xref="taxon:3544"
                      /clone_lib="MP"
                      /tissue_type="apical meristem and leaf primordia"
                      /dev_stage="6 weeks"
                      /note="3 d 500mM NaCl"
BASE COUNT      259 a    247 c    244 g    299 t      5 others
ORIGIN


  Query Match             49.4%;  Score 23.2;  DB 134;  Length 1054;
  Best Local Similarity   70.5%;  Pred. No. 42;
  Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      4 tataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
          ||| |||| ||| |  || |   || ||| ||||||| ||| ||
Db    597 TATGATGCAGCCTCTAATGCAATCTTATCAGCCGCAACCTTCGA 554


RESULT   5
AW977483/c
LOCUS       AW977483      705 bp    mRNA            EST       02-JUN-2000
DEFINITION  EST389592 MAGE resequences, MAGO Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW977483
VERSION     AW977483.1  GI:8168734
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 378
            Seq primer: Forward.
FEATURES              Location/Qualifiers
     source           1. .705
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="MAGE resequences, MAGO"
                      /note="Vector: pBluescriptSKm"
BASE COUNT      175 a    187 c    147 g    196 t
ORIGIN


  Query Match             47.7%;  Score 22.4;  DB 97;  Length 705;
  Best Local Similarity   72.5%;  Pred. No. 82;
  Matches   29;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      5 ataatgcggccgcatattcggcctgatcggccgcaagctt 44
          |||    || | ||  |  ||||||| ||||||||||||
Db     40 ATAGATAGGTCTCAGGTGGGGCCTGAGCGGCCGCAAGCTT 1


RESULT   6
BE062593/c
LOCUS       BE062593      528 bp    mRNA            EST       09-JUN-2000
DEFINITION  QV1-BT0260-281099-023-c01 BT0260 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE062593
VERSION     BE062593.1  GI:8407243
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
```

```
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV1-BT0260-281
            099-023-c01&t3=1999-10-28&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 9
            High quality sequence stop: 426.
FEATURES                Location/Qualifiers
     source             1. .528
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone_lib="BT0260"
                        /dev_stage="Adult"
                        /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                        SmaI; A mini-library was made by cloning products derived
                        from  ORESTES PCR (U.S. Letters Patent application No. 196
                        ,716 -  Ludwig Institute for Cancer Research) profiles
                        into  the pUC 18 vector. Reverse transcription of tissue
                        mRNA and cDNA amplification were performed under low
                        stringency conditions."
BASE COUNT      101 a    180 c    152 g     94 t      1 others
ORIGIN


  Query Match             47.2%;  Score 22.2;  DB 134;  Length 528;
  Best Local Similarity   77.1%;  Pred. No. 96;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      7 aatgcggccgcatattcggcctgatcggccgcaag 41
          ||| ||||||||  ||| ||  || |||| |||||
Db    332 AATTCGGCCGCACCTTCCGCGAGAACGGCAGCAAG 298


RESULT   7
AI514427/c
LOCUS       AI514427      625 bp    mRNA            EST       16-MAR-1999
DEFINITION  LD40932.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD40932 5prime similar to U13637: yl
            FBgn0004649 PID:g535346 SWISS-PROT:P98163, mRNA sequence.
ACCESSION   AI514427
VERSION     AI514427.1  GI:4418489
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            ,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 409  row: C  column: 8
            High quality sequence stop: 507.
FEATURES                Location/Qualifiers
     source             1. .625
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone="LD40932"
                        /clone_lib="LD Drosophila melanogaster embryo pOT2"
                        /sex="male and female"
                        /dev_stage="0 to 24 hours mixed stage embryonic"
                        /lab_host="XL1 Blue"
                        /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                        XhoI; Sized fractionated cDNAs were directly ligated into
                        pOT2. "
BASE COUNT      138 a    166 c    184 g    137 t
ORIGIN


  Query Match             47.2%;  Score 22.2;  DB 21;  Length 625;
  Best Local Similarity   69.8%;  Pred. No. 97;
  Matches   30;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      5 ataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
          | ||||||||||| || || ||       | ||| |||||||
Db    349 ACAATGCGGCCGCACATGCGTCCCACGAAACTGCACGCTTGGA 307


RESULT   8
AA952587
LOCUS       AA952587      267 bp    mRNA            EST       29-OCT-1998
DEFINITION  TENS1771 T. cruzi epimastigote normalized cDNA Library Trypanosoma
            cruzi cDNA clone 1771 5', mRNA sequence.
ACCESSION   AA952587
VERSION     AA952587.1  GI:3115683
KEYWORDS    EST.
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 267)
  AUTHORS   Verdun,R.E., Di Paolo,N.C., Urmenyi,T.P., Rondinelli,E., Frasch
            ,A.C.C. and Sanchez,D.O.
  TITLE     Gene discovery through expressed sequence tag sequencing in
            trypanosoma cruzi
  JOURNAL   Infect. Immun. 66 (11), 5393-5398 (1998)
  MEDLINE   99003155
COMMENT     Contact: Sanchez D.O.
            Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
            San Martin)
            Av. Gral Paz entre Albarellos y Constituyentes, INTI edificio 24
            cp(1650) San Martin, Prov. de BS AS. Argentina
            Tel: (54-1)752-9639 or (54-1)752-0021
            Fax: (54-1)752-0021 or (54-1)752-9639
            Email: dsanchez@inti.gov.ar
            Seq primer: T7.
FEATURES                Location/Qualifiers
     source             1. .267
                        /organism="Trypanosoma cruzi"
                        /strain="Cl-Brenner"
                        /db_xref="taxon:5693"
                        /clone="1771"
                        /clone_lib="T. cruzi epimastigote normalized cDNA Library"
                        /cell_type="epimastigote"
                        /note="cDNA library constructed with oligo dT primed
                        epimastigote mRNA and cloned in pt7t318D phagemid with
                        modified polylinker (PHARMACIA)"
BASE COUNT       66 a     79 c     72 g     46 t      4 others
ORIGIN


  Query Match             46.8%;  Score 22;  DB 14;  Length 267;
  Best Local Similarity   73.7%;  Pred. No. 1.1e+02;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      7 aatgcggccgcatattcggcctgatcggccgcaagctt 44
          | | | ||||||  | | ||| |  ||||||||||||
Db    200 AGTACTGCCGCAACTGCAGCCCGTGCGGCCGCAAGCTT 237


RESULT   9
AI667900
LOCUS       AI667900      313 bp    mRNA            EST       07-JUL-1999
DEFINITION  TENGO871 T. Cruzi epimastigote normalised cDNA Library Trypanosoma
            cruzi cDNA clone n442.r 5', mRNA sequence.
ACCESSION   AI667900
VERSION     AI667900.1  GI:4826272
```

```
KEYWORDS    EST.
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   Gonzalez Rey,E., Remisz,E., Delgado Garcia,A. and Gonzalez,A.
  TITLE     Characterization of ESTs from Trypanosoma cruzi epimastigotes
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Delgado Alberto
            Departamento de Biologia Molecular, Lab 303
            Instituto de Parasitologia y Biomedicina
            Consejo Superior de Investigaciones Cientificas C/ Ventanilla No 11
            , E-18001, Granada, Spain
            Tel: 34 958 805058
            Fax: 34 958 203323
            Similar to T.cruzi elongation factor 1-gamma (mammalian homologue)
            mRNA, gb|L17307|TRBEF1GHOM with Blastn.
            Seq primer: T7
            High quality sequence stop: 313.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Trypanosoma cruzi"
                     /strain="Cl - Brenner"
                     /db_xref="taxon:5693"
                     /clone="n442.r"
                     /clone_lib="T. Cruzi epimastigote normalised cDNA Library"
                     /cell_type="epimastigote"
                     /note="Site_1: EcoRI; Site_2: NotI; cDNA library
                     constructed with oligo dt primed  epimastigote mRNA and
                     cloned in pt7t318D  phagemid with modified polylinker"
BASE COUNT       72 a     91 c     83 g     67 t
ORIGIN


  Query Match             46.8%;  Score 22;  DB 23;  Length 313;
  Best Local Similarity   73.7%;  Pred. No. 1.1e+02;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       7 aatgcggccgcatattcggcctgatcggccgcaagctt 44
           | | | ||||||  | | ||| |  ||||||||||||
Db     218 AGTACTGCCGCAACTGCAGCCCGTGCGGCCGCAAGCTT 255


RESULT  10
AA063113/c
LOCUS       AA063113      447 bp    mRNA            EST       02-FEB-1997
DEFINITION  zf68e10.s1 Soares_pineal_gland_N3HPG Homo sapiens cDNA clone
            IMAGE:382122 3' similar to contains element LTR5 repetitive element
            ;, mRNA sequence.
ACCESSION   AA063113
VERSION     AA063113.1  GI:1556773
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 826   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 253.
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Homo sapiens"
                     /db_xref="GDB:1290379"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:382122"
                     /clone_lib="Soares_pineal_gland_N3HPG"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: pineal gland; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer
                     [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCGTTTTTTTTTTTTTTTTTT 3']
                     , double-stranded cDNA was size selected, ligated to Eco
                     RI adapters (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo. "
BASE COUNT      129 a    104 c    114 g     94 t      6 others
ORIGIN


  Query Match             46.8%;  Score 22;  DB 1;  Length 447;
  Best Local Similarity   70.0%;  Pred. No. 1.1e+02;
  Matches   28;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 agctataatgcggccgcatattcggcctgatcggccgcaa 40
           |||  ||||   || |||||  |||||||| ||  || |||
Db     252 AGCATTAATTTTGCTGCATANGCGGCCTGTTCTNCCACAA 213


RESULT  11
AI187524/c
LOCUS       AI187524      475 bp    mRNA            EST       02-OCT-2000
DEFINITION  EST316 Manduca sexta male antennae Uni-ZAP XR library Manduca sexta
            cDNA clone pMsmaD85 5' similar to cuticle protein, mRNA sequence.
ACCESSION   AI187524
VERSION     AI187524.1  GI:3738733
KEYWORDS    EST.
SOURCE      tobacco hornworm.
  ORGANISM  Manduca sexta
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia
            ; Sphingiodea; Sphingidae; Sphinginae; Manduca.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Robertson,H.M., Martos,R., Sears,C., Todres,E.Z., Walden,K.O. and
            Nardi,J.B.
  TITLE     Diversity of odourant binding proteins revealed by an expressed
            sequence tag project on male Manduca sexta moth antennae
  JOURNAL   Insect Mol. Biol. 8, 501-518 (1999)
COMMENT     Other_ESTs: EST315
            Contact: Robertson HM
            Department of Entomology
            University of Illinois at Urbana-Champaign
            505 S. Goodwin, Urbana, IL 61801, USA
            Tel: 217 333-0489
            Fax: 217 244 3499
            Email: hughrobe@uiuc.edu
            This insert is in reverse orientation.
            Insert Length: 1000   Std Error: 0.00
            Seq primer: T7
            High quality sequence stop: 300.
FEATURES             Location/Qualifiers
     source          1. .475
                     /organism="Manduca sexta"
                     /db_xref="taxon:7130"
                     /clone="pMsmaD85"
                     /clone_lib="Manduca sexta male antennae Uni-ZAP XR
                     library"
```

```
                /sex="male"
                /dev_stage="newly eclosed adults and pharate adults"
                /lab_host="XL1 Blue MRF and SOLR"
                /note="Organ: antennae; Vector: Uni-ZAP XR; Site_1: EcoRI;
                Site_2: XhoI; The library was prepared by Stratagene using
                oligo-T priming  and unidirectional cloning with an
                adaptor at the 5' end  (GGCACGAG) following the EcoRI
                site.  The mRNA was prepared from  antennae of late pupal
                and newly eclosed male moths.  Clones  were subcloned in
                vivo in mass into pBluescript maintained  in SOLR cells
                for DNA sequencing."
BASE COUNT      105 a    111 c    172 g     87 t
ORIGIN


  Query Match             46.8%;  Score 22;  DB 17;  Length 475;
  Best Local Similarity   73.7%;  Pred. No. 1.1e+02;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      1 agctataatgcggccgcatattcggcctgatcggccgc 38
          |||||   ||||  | ||||| ||||||  ||  |||||
Db    142 AGCTACGCTGCGCCAGCATACTCGGCCGCATATGCCGC 105


RESULT  12
BE535565
LOCUS       BE535565      892 bp    mRNA            EST       09-AUG-2000
DEFINITION  601059673F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3445938 5',
            mRNA sequence.
ACCESSION   BE535565
VERSION     BE535565.1  GI:9764210
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 892)
  AUTHORS   NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8417  row: b  column: 19
            High quality sequence stop: 612.
FEATURES              Location/Qualifiers
     source           1. .892
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:3445938"
                      /clone_lib="NIH_MGC_10"
                      /cell_line="MGC36"
                      /lab_host="DH10B"
                      /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                      Average insert size 1.5 kb. Library prepared by Life
                      Technologies."
BASE COUNT      189 a    280 c    255 g    168 t
ORIGIN


  Query Match             46.8%;  Score 22;  DB 109;  Length 892;
  Best Local Similarity   67.4%;  Pred. No. 1.2e+02;
  Matches   31;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      2 gctataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
```

```
          ||| ||| | ||| ||| || ||| |   | |||||   | |||||
Db    747 GCTCTAAAGGGGCAGCAGATGCGGGCATCTGGGCCGGGTGATTGGA 792


RESULT  13
AV624476
LOCUS       AV624476      436 bp    mRNA            EST       11-OCT-2000
DEFINITION  AV624476 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
            reinhardtii cDNA clone LC078c03_r 5', mRNA sequence.
ACCESSION   AV624476
VERSION     AV624476.1  GI:10773653
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 436)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of Expressed Sequence Tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. (2000) In press
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES              Location/Qualifiers
     source           1. .436
                      /organism="Chlamydomonas reinhardtii"
                      /strain="C9"
                      /db_xref="taxon:3055"
                      /clone="LC078c03_r"
                      /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
                      /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                      XhoI; The cDNA library was constructed from cells cultured
                      in a carbon stress acclimatized condition in which carbon
                      dioxide concentration in the bubbling gas was changed from
                      5% to 0.04%"
BASE COUNT       69 a    157 c    116 g     94 t
ORIGIN


  Query Match             46.4%;  Score 21.8;  DB 37;  Length 436;
  Best Local Similarity   78.8%;  Pred. No. 1.4e+02;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy     11 cggccgcatattcggcctgatcggccgcaagct 43
          |||||||| |||| || |||| |  ||| ||||
Db    402 CGGCCGCACATTCCGCATGATGGCGCGCCAGCT 434


RESULT  14
AZ342872/c
LOCUS       AZ342872      558 bp    DNA             GSS       29-SEP-2000
DEFINITION  1M0076E01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0076E01 F, DNA sequence.
ACCESSION   AZ342872
VERSION     AZ342872.1  GI:10420544
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 558)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
```

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0076 row: E column: 01
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 558.

```
FEATURES             Location/Qualifiers
     source          1. .558
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0076E01"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      126 a    149 c    135 g    148 t
ORIGIN
```

```
  Query Match             46.4%;  Score 21.8;  DB 170;  Length 558;
  Best Local Similarity   70.7%;  Pred. No. 1.4e+02;
  Matches   29;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 ctataatgcggccgcatattcggcctgatcggccgcaagct 43
           ||||||  || ||| | || |   |||| || ||| || |||
Db     362 CTATAAAGCAGCCACCTAGTTCCCCTGCTCAGCCCCAGGCT 322
```

```
RESULT  15
AW181098/c
LOCUS       AW181098      610 bp    mRNA            EST       17-NOV-1999
DEFINITION  MgA0062r MgA Library Mycosphaerella graminicola cDNA clone MgA0062
            3' similar to OMEGA-6 FATTY ACID DESATURASE, ENDOPLASMIC RETICULUM
            ISOZYME 2, mRNA sequence.
ACCESSION   AW181098
VERSION     AW181098.1  GI:6448332
KEYWORDS    EST.
SOURCE      Mycosphaerella graminicola.
  ORGANISM  Mycosphaerella graminicola
            Eukaryota; Fungi; Ascomycota; Ascomycota incertae sedis;
            Mycosphaerellaceae; Mycosphaerella.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Keon,J.P.R., Bailey,A.M. and Hargreaves,J.A.
  TITLE     A group of expressed cDNA sequences from the wheat fungal leaf
            blotch pathogen, Mycosphaerella graminicola
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Hargreaves JA
            Cell Biology Department
```

IACR-Long Ashton Research Station
Long Ashton, Bristol, BS41 9AF, UK
Tel: +44(0)1275 392181
Fax: +44(0)1275 394281
Email: john.hargreaves@bbsrc.ac.uk
Insert Length: 2200 Std Error: 0.00
Seq primer: M13 forward.

```
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Mycosphaerella graminicola"
                     /strain="Strit"
                     /db_xref="taxon:54734"
                     /clone="MgA0062"
                     /clone_lib="MgA Library"
                     /note="Vector: pSPORT1; Library constructed from cultures
                     utilizing ammomium ions as a source of nitrogen"
BASE COUNT      125 a    169 c    156 g    159 t      1 others
ORIGIN
```

```
  Query Match             46.4%;  Score 21.8;  DB 87;  Length 610;
  Best Local Similarity   78.8%;  Pred. No. 1.4e+02;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy      11 cggccgcatattcggcctgatcggccgcaagct 43
             || ||||  |||||||   ||||||||||| ||
Db     564 CGACCGCGAATTCGGCTACATCGGCCGCAACCT 532
```

```
Search completed: March  4, 2001, 11:46:05
Job time: 7055 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 12:07:04 ; Search time 96.53 Seconds
(without alignments)
182.908 Million cell updates/sec

Title: US-09-101-423B-8
Perfect score: 47
Sequence: 1 agctataatgcggccgcata.....tgatcggccgcaagcttgga 47

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*
1: /cgn2_2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	43	91.5	45	18	T72784	DNA tag. Syntheti
c	2	21.8	46.4	1503	19	V15518	Hamster oral papil
c	3	21.8	46.4	7647	19	V15519	Hamster oral papil
c	4	21.4	45.5	42	18	T61221	Primer HUCKN.NOT f
	5	21.2	45.1	500	20	X91235	T. gondii immunoge
	6	21.2	45.1	1818	21	Z46842	Semaphorin K1 gene
	7	21.2	45.1	3481	20	X21501	DNA sequence of so
c	8	20.8	44.3	8907	21	Z36926	Nucleotide sequenc
c	9	20.8	44.3	9181	20	X84323	Stealth virus nucl
	10	20.6	43.8	144460	21	Z93815	Olfactory receptor
	11	20.4	43.4	41	15	Q70114	Lambda INNER prime
	12	20.4	43.4	4260	9	N81768	Sequence encoding
	13	20.4	43.4	4380	9	N80222	Sequence of Mycoba
	14	20.4	43.4	4380	19	V05708	Mycobacterium tube
c	15	20.2	43.0	2885	20	X40055	Colon cancer assoc
	16	20	42.6	2008	21	A09260	Human alpha-2-delt
	17	20	42.6	3203	11	Q06760	Sequence of beta-g
	18	20	42.6	3598	21	A09261	Human alpha-2-delt
	19	20	42.6	3770	21	A09253	Human alpha-2-delt
	20	20	42.6	5059	21	Z36227	DNA encoding a per
c	21	19.8	42.1	1173	20	X91764	Porphorymonas ging
c	22	19.8	42.1	1218	20	X91630	Porphorymonas ging
	23	19.8	42.1	2178	20	Z12261	Neisseria gonorrho
	24	19.8	42.1	3415	17	T36481	Human integrin bet
	25	19.8	42.1	6727	18	T88014	Murine IL-5 cDNA g
	26	19.8	42.1	7673	19	V58229	Omega-cyclohexane
	27	19.6	41.7	3183	20	X26611	Trypanosoma cruzi
c	28	19.4	41.3	939	21	Z29251	Rifin 3193 gene.
c	29	19.4	41.3	2943	17	T16480	SA8 virus gB glyco
c	30	19.2	40.9	619	21	Z53637	Neisseria gonorrhe
c	31	19.2	40.9	834	21	Z53639	Neisseria meningit
c	32	19.2	40.9	1842	21	Z56349	Escherichia coli f
	33	19.2	40.9	1830121	17	T42063	Haemophilus influe
c	34	19	40.4	638	18	T45880	Human colon specif
c	35	19	40.4	638	19	V16668	Polynucleotide seq
	36	19	40.4	924	21	A12829	Pseudomonas cepaci
	37	19	40.4	3087	11	Q03634	Sequence for the 3
c	38	19	40.4	4394	13	Q21604	Alpha galactosidas
	39	19	40.4	5474	16	Q90251	Tuberous sclerosis
c	40	18.8	40.0	1095	21	Z54498	Neisseria meningit
	41	18.8	40.0	1229	20	Z42169	Human normal bladd
c	42	18.8	40.0	1797	20	X83965	Salmonella typhimu
	43	18.8	40.0	2302	21	A15621	Human heat shock p
	44	18.8	40.0	2302	21	A15622	Human heat shock p
	45	18.8	40.0	2465	18	T58086	Human heat shock p

ALIGNMENTS

RESULT 1
T72784/c
ID T72784 standard; DNA; 45 BP.
XX
AC T72784;
XX
DT 22-SEP-1997 (first entry)
XX
DE DNA tag.
XX
KW Metastasis-inducing DNA; Met-DNA; cancer; diagnosis; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 1..2
FT /*tag= a
FT /note= "5' single stranded overhang"
FT misc_feature 45
FT /*tag= b
FT /note= "5' overhang on complementary strand of 4
FT bases with sequence 5'-AGCT-3'"
XX
PN WO9725443-A1.
XX
PD 17-JUL-1997.
XX
PF 10-JAN-1997; 97WO-GB00074.
XX
PR 10-JAN-1996; 96GB-0000470.
XX
PA (UYLI-) UNIV LIVERPOOL.
XX
PI Barraclough BR, Rudland PS;
XX

```
DR   WPI; 1997-372878/34.
XX
PT   New isolated metastasis-inducing DNA - used to develop products to
PT   identify and treat patients at risk from metastatic tumours
XX
PS   Claim 6; Page 24; 38pp; English.
XX
CC   A DNA tag (T72784) having HindIII, SfiI, NotI and a defective
CC   HindIII site is used to tag human DNA from malignant, metastatic
CC   cancer cells in a novel method of screening and recovering
CC   metastasis-inducing DNA (Met-DNA).  This method involves
CC   transferring the human DNA into a cell line (pref. rat Rama 37)
CC   that produces only benign, non-metastasing tumours when injected
CC   into a syngeneic animal, injecting the transformed cells into a
CC   syngeneic animal, selecting those animals having metastasing
CC   tumours, and recovering the Met-DNA (see also T72785-90) from them.
CC   The human DNA is tagged to assist in its removal or insertion from
CC   or into a host or vector.  The tagging procedure overcomes the
CC   problem of identifying the inserted human DNA sequences in the rat
CC   genome of the transfected cells.
XX
SQ   Sequence 45 BP; 11 A; 13 C; 12 G; 9 T; 0 other;


  Query Match             91.5%;  Score 43;  DB 18;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 4e-08;
  Matches   43;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 ataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
           |||||||||||||||||||||||||||||||||||||||||||
Db      45 ATAATGCGGCCGCATATTCGGCCTGATCGGCCGCAAGCTTGGA 3


RESULT   2
V15518/c
ID   V15518 standard; DNA; 1503 BP.
XX
AC   V15518;
XX
DT   22-MAY-1998  (first entry)
XX
DE   Hamster oral papilloma virus L1 DNA.
XX
KW   L1 DNA; L1 protein; diagnosis; papilloma virus infectious disease;
KW   cancer; screening; antitumour agent; antisense treatment; ss.
XX
OS   Hamster oral papilloma virus.
XX
PN   JP10042875-A.
XX
PD   17-FEB-1998.
XX
PF   06-AUG-1996;   96JP-0207143.
XX
PR   06-AUG-1996;   96JP-0207143.
XX
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 1998-186869/17.
DR   P-PSDB; W47224.
XX
PT   DNA encoding hamster oral papilloma virus protein - useful for, e.g.
PT   diagnosing cancer related papilloma virus infections
XX
PS   Claim 10; Pages 13-14; 17pp; Japanese.
XX
CC   The present sequence is the hamster oral papilloma virus L1 DNA.
CC   The L1 protein and DNA can be used for the diagnosis of papilloma
CC   virus infectious diseases related to cancer, screening for
CC   antitumour agents and antisense treatment.
XX
SQ   Sequence 1503 BP; 434 A; 346 C; 337 G; 386 T; 0 other;


  Query Match             46.4%;  Score 21.8;  DB 19;  Length 1503;
  Best Local Similarity   70.7%;  Pred. No. 9.9;
  Matches   29;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       4 tataatgcggccgcatattcggcctgatcggccgcaagctt 44
             ||| ||||  | || | ||   ||||||| ||||||| | ||
Db     866 TATGATGCATCAGCCTTTTGATCCTGATCAGCCGCAGGTTT 826


RESULT   3
V15519/c
ID   V15519 standard; DNA; 7647 BP.
XX
AC   V15519;
XX
DT   22-MAY-1998  (first entry)
XX
DE   Hamster oral papilloma virus genomic DNA.
XX
KW   Diagnosis; papilloma virus infectious disease;
KW   cancer; screening; antitumour agent; antisense treatment; ss.
XX
OS   Hamster oral papilloma virus.
XX
PN   JP10042875-A.
XX
PD   17-FEB-1998.
XX
PF   06-AUG-1996;   96JP-0207143.
XX
PR   06-AUG-1996;   96JP-0207143.
XX
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 1998-186869/17.
XX
PT   DNA encoding hamster oral papilloma virus protein - useful for, e.g.
PT   diagnosing cancer related papilloma virus infections
XX
PS   Claim 11; Pages 14-16; 17pp; Japanese.
XX
CC   The present sequence is a hamster oral papilloma virus genomic
CC   DNA. The DNA can be used for the diagnosis of papilloma virus
CC   infectious diseases related to cancer, screening for antitumour
CC   agents and antisense treatment.
XX
SQ   Sequence 7647 BP; 2102 A; 1703 C; 1894 G; 1948 T; 0 other;


  Query Match             46.4%;  Score 21.8;  DB 19;  Length 7647;
  Best Local Similarity   70.7%;  Pred. No. 12;
  Matches   29;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       4 tataatgcggccgcatattcggcctgatcggccgcaagctt 44
             ||| ||||  | || | ||   ||||||| ||||||| | ||
Db    6514 TATGATGCATCAGCCTTTTGATCCTGATCAGCCGCAGGTTT 6474


RESULT   4
T61221/c
ID   T61221 standard; DNA; 42 BP.
XX
AC   T61221;
XX
DT   17-OCT-1997  (first entry)
XX
DE   Primer HUCKN.NOT for light chain constant region cDNA.
XX
KW   Primer; polymerase chain reaction; PCR; amplification; light chain;
KW   immunoglobulin; Ig; constant; region; production;
KW   recombinant; antibody; B cell; diagnosis; therapy; ss.
XX
```

OS Synthetic.
XX
PN DE19526546-A1.
XX
PD 23-JAN-1997.
XX
PF 20-JUL-1995; 95DE-1026546.
XX
PR 20-JUL-1995; 95DE-1026546.
XX
PA (OPEL/) OPELZ G.
XX
PI Terness P, Welschof M;
XX
DR WPI; 1997-088250/09.
XX
PT Prodn. of recombinant antibodies - by amplification, cloning and
PT expression of cDNA generated from B-cell mRNA
XX
PS Claim 5; Page 11; 11pp; German.
XX
CC The present sequence is a primer for the PCR amplification of the
CC cDNA encoding an immunoglobulin (Ig), light chain, constant region.
CC The primer can be used in a novel method for the production of
CC recombinant antibodies, comprising the selection of B cells from a
CC lymphocyte fraction, isolation of mRNA from individual B cells,
CC reverse transcription of the mRNA into cDNA, amplification of the
CC cDNA by PCR and cloning and expression of the cDNA. The recombinant
CC antibodies can be used for diagnosis and/or therapy, while the
CC method avoids the need for intermediate separation of light and
CC heavy chains and gene library screening.
XX
SQ Sequence 42 BP; 11 A; 11 C; 14 G; 6 T; 0 other;

Query Match 45.5%; Score 21.4; DB 18; Length 42;
Best Local Similarity 71.8%; Pred. No. 8.9;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy      7 aatgcggccgcatattcggcctgatcggccgcaagcttg 45
          | || ||| |||   || | ||   ||||||||||||||
Db     42 ACTGTGGCTGCACCATCTGTCTTCGCGGCCGCAAGCTTG 4
```

RESULT 5
X91235
ID X91235 standard; DNA; 500 BP.
XX
AC X91235;
XX
DT 24-SEP-1999 (first entry)
XX
DE T. gondii immunogenic protein encoding DNA.
XX
KW Immunogenic protein; Toxoplasma gondii protein; oocyst shedding; cat;
KW T. gondii infection; enteric apicomplexa oocyst; Cryptosporidium oocyst;
KW Toxoplasma oocyst; ss.
XX
OS Toxoplasma gondii.
XX
PN WO9932633-A1.
XX
PD 01-JUL-1999.
XX
PF 18-DEC-1998; 98WO-US27137.
XX
PR 19-DEC-1997; 97US-0994825.
XX
PA (HESK-) HESKA CORP.
XX
PI Lutz SB, Milhausen MJ, Ng RK;
XX
DR WPI; 1999-418930/35.
XX
PT New isolated Toxoplasma gondii nucleic acids used, e.g. to treat
PT infection caused by this microorganism
XX
PS Claim 4; Page 213; 381pp; English.
XX
CC The invention provides isolated Toxoplasma gondii nucleic acids that
CC encode immunogenic polypeptides. The T. gondii nucleic acid molecules,
CC immunogenic proteins and antibodies to the proteins can be used to
CC inhibit T. gondii oocyst shedding in a cat due to infection with
CC T. gondii. They can be used for preventing T. gondii infection and for
CC preventing the spread of T. gondii infection. They can also be used for
CC detecting T. gondii infection. The detection method can be used to detect
CC parasite cysts or oocysts in feces, e.g. from enteric apicomplexa oocysts
CC such as Cryptosporidium oocysts and Toxoplasma oocysts.
XX
SQ Sequence 500 BP; 118 A; 118 C; 131 G; 133 T; 0 other;

Query Match 45.1%; Score 21.2; DB 20; Length 500;
Best Local Similarity 76.5%; Pred. No. 15;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy     11 cggccgcatattcggcctgatcggccgcaagctt 44
            | ||||||  ||||||| ||||||| | ||  ||
Db    440 ccgccgcaagttcggcccgatcggctggaacatt 473
```

RESULT 6
Z46842
ID Z46842 standard; cDNA; 1818 BP.
XX
AC Z46842;
XX
DT 10-APR-2000 (first entry)
XX
DE Semaphorin K1 gene related sequence.
XX
KW Semaphorin K1; cellular physiology; neurite outgrowth; neuron; human;
KW immunogen; pharmaceutical; ss.
XX
OS Unidentified.
XX
PN JP11341988-A.
XX
PD 14-DEC-1999.
XX
PF 11-MAR-1999; 99JP-0065672.
XX
PR 11-MAR-1998; 98US-0041236.
XX
PA (EXEL-) EXELIXIS PHARM INC.
XX
DR WPI; 2000-109378/10.
DR P-PSDB; Y56855.
XX
PT New semaphorin polypeptides, useful cell physiology modulators and
PT immunogens -
XX
PS Disclosure; Page 17-20; 57pp; Japanese.
XX
CC The invention provided isolated human semaphorin K1 polypeptides. The
CC polypeptides, or nucleic acids encoding them, can be used to modulate
CC cellular physiology by modulating semaphorin K1 activity, e.g. semaphorin
CC K1 polypeptide fragments or antisense nucleic acids can be used to
CC enhance neurite outgrowth from damaged neurons. The polypeptides can also
CC be used as immunogens, reagents for isolating other semaphorins, or as
CC reagents for screening chemical libraries for lead pharmaceutical agents.
CC The nucleic acids can also be used as probes and primers for diagnostic
CC purposes.
XX
SQ Sequence 1818 BP; 543 A; 409 C; 359 G; 507 T; 0 other;

```
  Query Match             45.1%;  Score 21.2;  DB 21;  Length 1818;
  Best Local Similarity   69.0%;  Pred. No. 17;
  Matches   29;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      4 tataatgcggccgcatattcggcctgatcggccgcaagcttg 45
          |||||  | ||  ||||  | | | ||| ||| || ||||||
Db    312 tataactctgcttcataaccagacagatgggctgctagcttg 353


RESULT   7
X21501
ID   X21501 standard; DNA; 3481 BP.
XX
AC   X21501;
XX
DT   21-MAY-1999  (first entry)
XX
DE   DNA sequence of sorbitol dehydrogenase (SLDH) and ORF2 genes.
XX
KW   Sorbitol dehydrogenase; SLDH; open reading frame; ORF2; L-sorbose;
KW   D-sorbitol; vitamin C; ds.
XX
OS   Gluconobacter suboxydans.
XX
FH   Key             Location/Qualifiers
FT   RBS             177..182
FT                   /*tag=  a
FT                   /note= "putative Shine-Dalgarno (SD) sequence for
FT                         ORF2 gene"
FT   CDS             192..572
FT                   /*tag=  b
FT                   /product= "ORF2 gene product"
FT                   /note= "corresponding protein sequence shown in W95020"
FT   RBS             558..564
FT                   /*tag=  c
FT                   /note= "putative Shine-Dalgarno (SD) sequence for
FT                         SLDH gene"
FT   CDS             572..2794
FT                   /*tag=  d
FT                   /product= "SLDH gene product"
FT                   /note= "corresponding protein sequence shown in W95019"
FT   sig_peptide     572..643
FT                   /*tag=  e
FT                   /note= "signal sequence for SLDH"
FT   mat_peptide     644..2791
FT                   /*tag=  f
FT                   /note= "mature SLDH protein sequence"
FT   repeat_region   684..704
FT                   /*tag=  g
FT                   /rpt_type= inverted
FT                   /note= "inverted repeat sequence IR1 as possible
FT                         transcription terminator for ORF2 gene as
FT                         indicated in the specification"
FT   repeat_unit     684..693
FT                   /*tag=  h
FT   repeat_region   2803..2892
FT                   /*tag=  i
FT                   /rpt_type= inverted
FT                   /note= "inverted repeat sequence IR2 as possible
FT                         transcription terminator for SLDH gene as
FT                         indicated in the specification"
FT   repeat_unit     2803..2833
FT                   /*tag=  j
XX
PN   EP897984-A2.
XX
PD   24-FEB-1999.
XX
PF   13-AUG-1998;   98EP-0115231.
XX
PR   21-AUG-1997;   97EP-0114432.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Hosino T,  Miyazaki T,  Ojima S,  Shinjoh M,  Tomiyama N;
XX
DR   WPI; 1999-134646/12.
DR   P-PSDB; W95019, W95020.
XX
PT   New D-sorbitol dehydrogenase gene and recombinant protein - useful
PT   for production of L-sorbose, an intermediate in vitamin C production
XX
PS   Claim 4; Fig 3A-D; 39pp; English.
XX
CC   This represents the DNA sequence of the SLDH gene encoding a protein with
CC   sorbitol dehydrogenase (SLDH) activity. The DNA also encodes an open
CC   reading frame (ORF2) product upstream of the SLDH open reading frame,
CC   needed for SLDH activity in vivo. Host cells transformed by a vector
CC   comprising the SLDH DNA sequence are used for the recombinant expression
CC   of the sorbitol dehydrogenase. Recombinant D-sorbitol dehydrogenase is
CC   useful for producing L-sorbose from D-sorbitol. L-sorbose is an important
CC   intermediate in vitamin C production.
XX
SQ   Sequence 3481 BP; 674 A; 1018 C; 1001 G; 788 T; 0 other;


  Query Match             45.1%;  Score 21.2;  DB 20;  Length 3481;
  Best Local Similarity   76.5%;  Pred. No. 19;
  Matches   26;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy     13 gccgcatattcggcctgatcggccgcaagcttgg 46
          || || |||||   || || ||||||| ||||||
Db    355 gcggcgtattcatgctcatgggccgcacgcttgg 388


RESULT   8
Z36926/c
ID   Z36926 standard; DNA; 8907 BP.
XX
AC   Z36926;
XX
DT   13-MAR-2000  (first entry)
XX
DE   Nucleotide sequence of the genome of Stealth virus clone 313.
XX
KW   Prototype Stealth virus clone; atypically structured virus;
KW   vacuolating cytopathic effect; cytopathic virus; virus detection;
KW   malignancy; multiple myeloma; lymphoma; brain tumour; breast cancer;
KW   salivary gland tumour; Alzheimer's disease; Parkinson's disease;
KW   spongiform encephalopathy; multiple sclerosis; schizophrenia;
KW   manic depression; major depression; personality disorder; autism;
KW   Rett's syndrome; attention deficit; oppositional defiance; aggression;
KW   anorexia nervosa; bulimia; multi-system illness; virus infection; ss.
XX
OS   Stealth virus.
XX
PN   WO9960101-A1.
XX
PD   25-NOV-1999.
XX
PF   19-MAY-1999;   99WO-US11185.
XX
PR   19-MAY-1998;   98US-0081708.
XX
PA   (MART/) MARTIN W J.
XX
PI   Martin WJ;
XX
DR   WPI; 2000-072436/06.
XX
PT   Isolated viruses, used to develop products for detection, prevention
PT   and treatment of stealth virus infections  -
XX
PS   Disclosure; Page 34-37; 50pp; English.
XX
```

```
CC   Z36925-30 represent the nucleotide sequences of prototype Stealth virus
CC   clones. The sequences illustrate the type of sequences that can be used
CC   to generate probes and to predict antigenic and biologically active
CC   products obtainable from a viral isolate, in the methods of the
CC   invention. The specification describes tissue culture, serological and
CC   molecular based methods to detect atypically structured viruses,
CC   such as Stealth viruses, which are able to induce a vacuolating
CC   cytopathic effect (CPE) in tissue culture, and are distinguishable
CC   from known cytopathic viruses, by appearance progression and/or host
CC   range, or by serological, electron microscopic and/or molecular markers.
CC   The products can be used for detecting viruses in patients with
CC   diseases such as a malignancy, e.g. multiple myeloma, lymphoma, brain
CC   tumours, breast cancer, salivary gland tumours, Alzheimer's disease,
CC   Parkinson's disease, spongiform encephalopathy, multiple sclerosis,
CC   schizophrenia, manic depression, major depression, personality
CC   disorders, autism, Rett's syndrome, attention deficit, oppositional
CC   defiance, aggression, anorexia nervosa, bulimia, a multi-system
CC   illness, an animal illness or an illness in a domestic dog or cat.
CC   The products can be used for detecting Stealth viruses in biological
CC   products such as blood products and foods. The products can also be
CC   used for developing agents for treating or preventing virus infections.
XX
SQ   Sequence 8907 BP; 2257 A; 2447 C; 2165 G; 2023 T; 15 other;


  Query Match             44.3%;  Score 20.8;  DB 21;  Length 8907;
  Best Local Similarity   73.5%;  Pred. No. 30;
  Matches   25;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      12 ggccgcatattcggcctgatcggccgcaagcttg 45
           ||| |:||  |||||||||||||||| | |   ||
Db    6047 GGCTGYATGGTCGGCCTGATCGGCCCCGATGGTG 6014


RESULT   9
X84323/c
ID   X84323 standard; DNA; 9181 BP.
XX
AC   X84323;
XX
DT   08-SEP-1999  (first entry)
XX
DE   Stealth virus nucleic acid clone, SEQ ID NO: 15.
XX
KW   Stealth virus; detection; diagnosis; infection; ss.
XX
OS   Stealth virus.
XX
FH   Key             Location/Qualifiers
FT   misc_difference 8920
FT                   /*tag=  a
FT                   /note= "this nucleotide is represented as a * in the
FT                          specification, and is included to maintain the
FT                          base numbering given in the specification"
FT   misc_difference 8929
FT                   /*tag=  b
FT                   /note= "this nucleotide is represented as a * in the
FT                          specification, and is included to maintain the
FT                          base numbering given in the specification"
FT   misc_difference 8941
FT                   /*tag=  c
FT                   /note= "this nucleotide is represented as a * in the
FT                          specification, and is included to maintain the
FT                          base numbering given in the specification"
FT   misc_feature    8945..8947
FT                   /*tag=  d
FT                   /note= "these nucleotides are represented as * in the
FT                          specification, and are included to maintain the
FT                          base numbering given in the specification"
FT   misc_feature    8952..8954
FT                   /*tag=  e
FT                   /note= "these nucleotides are represented as * in the
FT                          specification, and are included to maintain the
FT                          base numbering given in the specification"
FT   misc_difference 8979
FT                   /*tag=  f
FT                   /note= "this nucleotide is represented as a * in the
FT                          specification, and is included to maintain the
FT                          base numbering given in the specification"
FT   misc_difference 9097
FT                   /*tag=  g
FT                   /note= "this nucleotide is represented as a * in the
FT                          specification, and is included to maintain the
FT                          base numbering given in the specification"
XX
PN   WO9934019-A1.
XX
PD   08-JUL-1999.
XX
PF   30-DEC-1998;   98WO-US27744.
XX
PR   30-DEC-1997;   97US-0001184.
XX
PA   (MART/) MARTIN W J.
XX
PI   Martin WJ;
XX
DR   WPI; 1999-405521/34.
XX
PT   Novel strains of stealth virus
XX
PS   Claim 19; Page 49-53; 95pp; English.
XX
CC   This sequence represents a Stealth virus nucleic acid clone. The
CC   invention relates to a method of detecting and characterising a stealth
CC   virus by reacting a sample suspected of containing a stealth virus with a
CC   probe from a known stealth virus and sequencing the resultant isolated
CC   nucleotide. The method comprises the steps of: (a) isolating DNA or RNA
CC   from a sample suspected of containing a stealth virus, e.g. a culture of
CC   cells showing a viral cytopathic effect; (b) testing the reactivity of
CC   the isolated DNA or RNA with a molecular probe that contains at least 18
CC   or more contiguous nucleotides identical to sequence previously
CC   identified from a stealth virus; and, optionally (c) sequencing the
CC   isolated DNA or RNA molecules that react with the probe. The method is
CC   used to detect stealth virus in a biological product, food or in the
CC   environment. The method is also used to evaluate agents for their
CC   inhibitory or stimulatory effects on stealth virus replication and to
CC   determine capacity of the virus to recombine with and potentially alter
CC   the nucleic acid sequences of a cell or bacterium.
XX
SQ   Sequence 9181 BP; 2287 A; 2465 C; 2173 G; 2050 T; 206 other;


  Query Match             44.3%;  Score 20.8;  DB 20;  Length 9181;
  Best Local Similarity   73.5%;  Pred. No. 31;
  Matches   25;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      12 ggccgcatattcggcctgatcggccgcaagcttg 45
           ||| |:||  |||||||||||||||| | |   ||
Db    6053 GGCTGYATGGTCGGCCTGATCGGCCCCGATGGTG 6020


RESULT  10
Z93815
ID   Z93815 standard; DNA; 144460 BP.
XX
AC   Z93815;
XX
DT   16-AUG-2000  (first entry)
XX
DE   Olfactory receptor operon.
XX
KW   Olfactory receptor protein; olfactory; smell; receptor; binding;
KW   detection; screening; genotyping; biallelic marker; human; ss.
XX
OS   Homo sapiens.
```

```
XX
FH   Key             Location/Qualifiers
FT   CDS             2406..2600
FT                   /*tag= a
FT                   /label= ORF1
FT   CDS             9711..10658
FT                   /*tag= b
FT                   /label= ORF2
FT   CDS             24851..25369
FT                   /*tag= c
FT                   /label= ORF3
FT   CDS             45714..46661
FT                   /*tag= d
FT                   /label= ORF4
FT   CDS             80198..81115
FT                   /*tag= e
FT                   /label= ORF 5
FT   CDS             96291..96902
FT                   /*tag= f
FT                   /label= ORF6
FT   CDS             110758..111564
FT                   /*tag= g
FT                   /label= ORF7
FT   CDS             122525..122887
FT                   /*tag= h
FT                   /label= ORF8
FT   CDS             132454..133389
FT                   /*tag= i
FT                   /label= ORF9
FT   CDS             143398..143577
FT                   /*tag= j
FT                   /label= ORF10
FT   misc_feature    complement (113644..114063)
FT                   /*tag= k
FT                   /note= "Ubiquitin 1 pseudogene complement"
FT   misc_feature    complement (127489..127854)
FT                   /*tag= l
FT                   /note= "Ubiquitin 2 pseudogene complement"
FT   allele          7521
FT                   /*tag= m
FT                   /note= "Polymorphic base G or T"
FT   allele          8192
FT                   /*tag= n
FT                   /note= "Polymorphic base G or T"
FT   allele          14483
FT                   /*tag= o
FT                   /note= "Polymorphic base A or T"
FT   allele          19625
FT                   /*tag= p
FT                   /note= "Polymorphic base C or T"
FT   allele          20583
FT                   /*tag= q
FT                   /note= "Polymorphic base A or G"
FT   allele          76947
FT                   /*tag= r
FT                   /note= "Polymorphic base A or C"
FT   allele          91088
FT                   /*tag= s
FT                   /note= "Polymorphic base G or C"
FT   allele          91138
FT                   /*tag= t
FT                   /note= "Polymorphic base A or G"
FT   allele          91187
FT                   /*tag= u
FT                   /note= "Polymorphic base A or G"
FT   allele          133998
FT                   /*tag= v
FT                   /note= "Polymorphic base A or G"
FT   allele          140066
FT                   /*tag= w
FT                   /note= "Polymorphic base C or T"
FT   allele          141176
FT                   /*tag= x
FT                   /note= "Polymorphic base A or G"
FT   allele          144033
FT                   /*tag= y
FT                   /note= "Polymorphic base A or G"
XX
PN   WO200021985-A2.
XX
PD   20-APR-2000.
XX
PF   13-OCT-1999;   99WO-IB01729.
XX
PR   14-OCT-1998;   98US-0104299.
XX
PA   (GEST ) GENSET.
XX
PI   Bougueleret L,  Malekzadeh K;
XX
DR   WPI; 2000-317933/27.
XX
PT   New nucleic acids encoding ten different olfactory receptor proteins
PT   and their biallelic markers, are useful in genetic analysis and in
PT   screening for compounds which bind to the receptor proteins
XX
PS   Claim 1; Page 103-141; 155pp; English.
XX
CC   Ten new olfactory receptor proteins and their biallelic markers have
CC   been described. The sequences encoding these receptor proteins and
CC   which contain the biallelic markers can be used for genotyping. The
CC   olfactory receptor proteins can be used to screen for substances
CC   which bind to them. See GENESEQ records Z93816-25 and Y83386-95.
XX
SQ   Sequence 144460 BP; 46068 A; 27088 C; 26615 G; 44676 T; 13 other;


  Query Match             43.8%;  Score 20.6;  DB 21;  Length 144460;
  Best Local Similarity   67.4%;  Pred. No. 52;
  Matches   29;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       2 gctataatgcggccgcatattcggcctgatcggccgcaagctt 44
           ||||||| ||| |||| | |||||| |||| | |||
Db  113793 gctataagatagcctcataaactgcctgaaaggccttattctt 113835


RESULT  11
Q70114
ID   Q70114 standard; DNA; 41 BP.
XX
AC   Q70114;
XX
DT   10-MAR-1995  (first entry)
XX
DE   Lambda INNER primer.
XX
KW   Enterokinase; EK; heavy chain; light chain; catalytic domain;
KW   digestive disorder; cleavage; fusion protein; trypsinogen;
KW   trypsin; enzyme; PACE gene; ss.
XX
OS   Synthetic.
XX
PN   WO9416083-A.
XX
PD   21-JUL-1994.
XX
PF   13-JAN-1994;   94WO-US00616.
XX
PR   15-JAN-1993;   93US-0005944.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Lavallie ER;
XX
DR   WPI; 1994-249229/30.
XX
```

PT New nucleic acid encoding enterokinase activity - and related
PT vectors, host cells, expression products and antibodies are
PT useful in treating digestive disorders and for cleaving fusion
PT proteins
XX
PS Example 6; Page 19; 50pp; English.
XX
CC Nested oligonucleotide primers were synthesised which were
CC complementary to the lambda DNA sequence adjacent to the cloning site
CC for the cDNA insertions (see Q70113-14). In addition, primers were
CC designed which were complementary to the plus strand of the most 5'
CC region of the EK coding sequence (see Q70115-16). Subcloning and
CC sequencing provided a 116 amino acid EK heavy chain. This protein
CC contained regions corresp. to tryptic peptides (see R57295-9).
XX
SQ Sequence 41 BP; 12 A; 10 C; 11 G; 8 T; 0 other;

Query Match 43.4%; Score 20.4; DB 15; Length 41;
Best Local Similarity 71.1%; Pred. No. 22;
Matches 27; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
Qy      5 ataatgcggccgcatattcggcctgatcggccgcaagc 42
          | ||||||||||||  ||| ||||| |      |||||
Db      4 agaatgcggccgcaagttcagcctggttaagtccaagc 41
```

RESULT 12
N81768
ID N81768 standard; DNA; 4260 BP.
XX
AC N81768;
XX
DT 29-DEC-1990 (first entry)
XX
DE Sequence encoding Mycobacterium tuberculosis 540 and 517 AA residue
DE proteins.
XX
KW Diagnosis; assay; M.bovis; vaccine; ds.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT CDS 252..1874
FT /*tag= a
FT /label=540 AA protein
FT /note="P81351"
FT CDS complement (3948..2395)
FT /*tag= b
FT /label=517 AA protein
FT /note="P81868"
XX
PN WO8806591-A.
XX
PD 07-SEP-1988.
XX
PF 25-FEB-1988; 88WO-US00598.
XX
PR 24-FEB-1988; 88US-0159667.
PR 06-FEB-1987; 87US-0019529.
XX
PA (SCRI-) SCRIPPS CLINIC & RE.
XX
PI Shinnick T, Houghten R;
XX
DR WPI; 1988-271136/38.
DR P-PSDB; P81351, P81868.
XX
PT Recombinant mycobacterial peptide(s) -
PT used in assays for diagnosis of infection, for producing
PT vaccines and for producing antibodies
XX
PS Disclosure; Fig 2a-2d; 116pp; English.
XX
CC An isolated DNA molecule that consists essentially of the nucleotide
CC sequence that corresponds to the sequence represented by position 3950 to
CC about 2390 and from position 3948 through position 2398 of N81768 is
CC claimed. Also claimed is a peptide sequence that consists of a 5-40 AA
CC residue sequence that corresponds to a sequence of the 540 AA residue
CC protein (P81351) or the 517 AA residue protein (P81868) coded for by the
CC DNA sequence. The proteins can be used for determining previous
CC immunological exposure of a mammal to M.tuberculosis or M.bovis and
CC for producing a vaccine.
XX
SQ Sequence 4260 BP; 733 A; 1332 C; 1481 G; 714 T; 0 other;

Query Match 43.4%; Score 20.4; DB 9; Length 4260;
Best Local Similarity 80.0%; Pred. No. 39;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy      9 tgcggccgcatattcggcctgatcggccgc 38
          |||||||||  | ||| |||||  | |||||
Db   3481 tgcggccgcttgttccgcctgcgctgccgc 3510
```

RESULT 13
N80222
ID N80222 standard; DNA; 4380 BP.
XX
AC N80222;
XX
DT 19-MAR-1991 (first entry)
XX
DE Sequence of Mycobacterium tuberculosis DNA contg. gene encoding 65
DE protein.
XX
KW Antigen; vaccine; ds.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT CDS 192..1874
FT /*tag= a
FT CDS complement (2398..4101)
FT /*tag= b
XX
PN WO8805823-A.
XX
PD 11-AUG-1988.
XX
PF 01-FEB-1988; 88WO-US00281.
XX
PR 02-FEB-1987; 87US-0010007.
XX
PA (WHIT-) WHITEHEAD INST BIOM.
XX
PI Husson RN, Young RA, Shinnick TM;
XX
DR WPI; 1988-235175/33.
DR P-PSDB; P80215, P80216
XX
PT Genes encoding Mycobacterium tuberculosis protein antigens -
PT useful for developing reagents for diagnosis, prevention and
PT treatment of tuberculosis
XX
PS Claim 12; Fig 8; 82pp; English.
XX
CC The gene was isolated by probing a lambda gt11 expression library of
CC M.tuberculosis DNA with monoclonal antibodies directed against
CC M.tuberculosis-specific antigens. The 19kD, 71kD and the 65kD proteins
CC and genes are claimed, and so is a vaccine comprising DNA encoding
CC M.tuberculosis protein in a recombinant vaccine vector. P80216 is
CC encoded on the complementary strand.
XX
SQ Sequence 4380 BP; 757 A; 1373 C; 1512 G; 738 T; 0 other;

```
  Query Match             43.4%;  Score 20.4;  DB 9;  Length 4380;
  Best Local Similarity   80.0%;  Pred. No. 40;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      9 tgcggccgcatattcggcctgatcggccgc 38
          ||||||||| | ||| ||||| | |||||
Db   3601 tgcggccgcttgttccgcctgcgctgccgc 3630


RESULT  14
V05708
ID   V05708 standard; DNA; 4380 BP.
XX
AC   V05708;
XX
DT   22-JUN-1998  (first entry)
XX
DE   Mycobacterium tuberculosis 65 kDa heat shock protein gene.
XX
KW   Heat shock protein; Mt Hsp65; autoimmune disease; immunotherapy;
KW   gene therapy; rheumatoid arthritis; multiple sclerosis; ds.
XX
OS   Mycobacterium tuberculosis.
XX
FH   Key             Location/Qualifiers
FT   CDS             252..1874
FT                   /*tag=  a
FT                   /product= 65 kDa heat shock protein
XX
PN   WO9746253-A2.
XX
PD   11-DEC-1997.
XX
PF   03-JUN-1997;   97WO-US09427.
XX
PR   03-JUN-1997;   97US-0019100.
PR   03-JUN-1996;   96US-0019100.
XX
PA   (AURA-) AURAGEN INC.
XX
PI   Haynes JR,  Prayaga SK,  Ramshaw IA;
XX
DR   WPI; 1998-041892/04.
DR   P-PSDB; W44702.
XX
PT   Treatment of autoimmune diseases - by administering
PT   autoantigen-coated particles or autoantigen-encoding nucleic acid
PT   construct
XX
PS   Example 2; Page 55-59; 72pp; English.
XX
CC   This DNA sequence encodes the 65 kDa heat shock protein (see
CC   W44702), designated Mt Hsp65, of Mycobacterium tuberculosis.  This
CC   protein cross-reacts with a component of articular cartilage, human
CC   Hsp60, that is up-regulated in the joints of arthritic patients.  A
CC   claimed method for treating or preventing an autoimmune disease in
CC   a mammal comprises: (a) providing a particle coated with an antigen
CC   against which an immune response is mounted in the autoimmune
CC   disease; (b) delivering the particle into the recipient cell of the
CC   mammal; and (c) repeating step (b) until either a reduction in a
CC   cytotoxic immune response or a desensitizing immune response is
CC   induced in the mammal.  Alternatively, step (a) comprises providing
CC   a nucleic acid construct comprising a coding sequence for the
CC   antigen, operably linked to control elements such that the coding
CC   sequence can be transcribed and translated in a recipient cell, and
CC   delivering the construct to the recipient cell using a gene gun.
CC   The antigen of step (a) is selected from collagen, Mt Hsp65,
CC   myelin basic protein, myelin oligodendrocyte glycoprotein,
CC   proteolipid protein, and epitopes thereof.  These antigens mitigate
CC   cytotoxic responses and elicit antigen desensitisation.  The method
CC   is used especially for treating rheumatoid arthritis or multiple
CC   sclerosis.  It represents a novel use for the known Mt Hsp65 gene.
XX
SQ   Sequence 4380 BP; 757 A; 1371 C; 1514 G; 738 T; 0 other;


  Query Match             43.4%;  Score 20.4;  DB 19;  Length 4380;
  Best Local Similarity   80.0%;  Pred. No. 40;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      9 tgcggccgcatattcggcctgatcggccgc 38
          ||||||||| | ||| ||||| | |||||
Db   3601 tgcggccgcttgttccgcctgcgctgccgc 3630


RESULT  15
X40055/c
ID   X40055 standard; DNA; 2885 BP.
XX
AC   X40055;
XX
DT   02-JUL-1999  (first entry)
XX
DE   Colon cancer associated gene.
XX
KW   Cancer associated antigen; diagnosis; research; treatment; human;
KW   breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW   prostate cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9904265-A2.
XX
PD   28-JAN-1999.
XX
PF   15-JUL-1998;   98WO-US14679.
XX
PR   22-JUN-1998;   98US-0102322.
PR   17-JUL-1997;   97US-0896164.
PR   10-OCT-1997;   97US-0061599.
PR   10-OCT-1997;   97US-0061765.
PR   10-OCT-1997;   97US-0948705.
PR   11-OCT-1997;   97GB-0021697.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Chen Y,  Gout I,  Gure A,  O'Hare M,  Obata Y,  Old LJ;
PI   Pfreundschuh M,  Sahin U,  Scanlan MJ,  Stockert E;
PI   Tureci O;
XX
DR   WPI; 1999-132448/11.
XX
PT   New isolated cancer associated nucleic acids and polypeptides -
PT   isolated using sera from cancer patients, used to develop products
PT   for the diagnosis, monitoring or treatment of cancers
XX
PS   Claim 67; Page 651-652; 787pp; English.
XX
CC   The invention relates to a method for diagnosing a disorder characterised
CC   by expression of a human cancer associated antigen precursor coded for by
CC   a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC   biological sample isolated from a subject with an agent that specifically
CC   binds to the NAM, an expression product or a fragment of an expression
CC   product complexed with an HLA molecule; and (b) determining the
CC   interaction between the agent and the NAM or the expression product as a
CC   determination of the disorder. The products and methods can be used in
CC   the diagnosis, monitoring, research, or treatment of conditions
CC   characterised by the expression of various cancer associated antigens.
CC   The invention provides nucleic acid sequences and encoded polypeptides
CC   which are cancer associated antigen precursors expressed in human breast
CC   cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC   lung cancer.
XX
SQ   Sequence 2885 BP; 626 A; 901 C; 838 G; 520 T; 0 other;
```

Query Match 43.0%; Score 20.2; DB 20; Length 2885;
Best Local Similarity 75.8%; Pred. No. 45;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy      10 gcggccgcatattcggcctgatcggccgcaagc 42
           || ||||||  || |||||| |||||| | | |
Db    2747 GCTGCCGCAGCTTGGGCCTGTTCGGCCCCCACC 2715
```

Search completed: March 4, 2001, 12:07:09
Job time: 5664 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 12:05:17 ; Search time 106.76 Seconds
(without alignments)
70.949 Million cell updates/sec

Title: US-09-101-423B-8
Perfect score: 47
Sequence: 1 agctataatgcggccgcata.....tgatcggccgcaagcttgga 47

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		21.2	45.1	3481	3	US-09-136-251-1	Sequence 1, Appli
2		20.4	43.4	41	1	US-08-200-900A-26	Sequence 26, Appl
3		20.4	43.4	41	4	PCT-US94-00616-26	Sequence 26, Appl
4	c	20.2	43.0	2885	3	US-08-948-705-2	Sequence 2, Appli
5		19.8	42.1	3415	1	US-08-054-077C-1	Sequence 1, Appli
6		19.8	42.1	6727	3	US-08-629-643A-5	Sequence 5, Appli
7		19.8	42.1	6727	3	US-09-280-799-1	Sequence 1, Appli
8	c	19.4	41.3	2943	1	US-08-042-747A-7	Sequence 7, Appli
9	c	19	40.4	638	1	US-08-469-667-1	Sequence 1, Appli
10	c	19	40.4	638	4	PCT-US95-07289-1	Sequence 1, Appli
11	c	18.8	40.0	1797	2	US-08-853-659A-28	Sequence 28, Appl
12	c	18.8	40.0	8967	2	US-08-853-659A-6	Sequence 6, Appli
13		18.8	40.0	8967	2	US-08-853-659A-9	Sequence 9, Appli
14	c	18.8	40.0	8967	2	US-08-853-659A-64	Sequence 64, Appl
15		18.8	40.0	8967	2	US-08-853-659A-67	Sequence 67, Appl
16	c	18.8	40.0	24701	2	US-08-853-659A-2	Sequence 2, Appli
17		18.8	40.0	24701	2	US-08-853-659A-3	Sequence 3, Appli
18	c	18.8	40.0	24701	2	US-08-853-659A-60	Sequence 60, Appl
19		18.8	40.0	24701	2	US-08-853-659A-61	Sequence 61, Appl
20	c	18.8	40.0	38506	3	US-09-320-878-19	Sequence 19, Appl
21		18.6	39.6	1708	4	PCT-US95-02481-3	Sequence 3, Appli
22		18.6	39.6	8051	2	US-08-576-626A-2	Sequence 2, Appli
23	c	18.4	39.1	68750	3	US-09-335-409-1	Sequence 1, Appli
24	c	18.2	38.7	1374	1	US-08-278-630A-9	Sequence 9, Appli
25	c	18.2	38.7	1424	1	US-08-403-634-3	Sequence 3, Appli
26	c	18.2	38.7	1424	3	US-08-913-441B-3	Sequence 3, Appli
27	c	18.2	38.7	1584	3	US-08-907-229-1	Sequence 1, Appli
28		18.2	38.7	1867	2	US-08-516-801-1	Sequence 1, Appli
29		18.2	38.7	1867	4	PCT-US95-06683-1	Sequence 1, Appli
30	c	18.2	38.7	1965	1	US-08-258-420-9	Sequence 9, Appli
31	c	18.2	38.7	2001	3	US-08-850-961-13	Sequence 13, Appl
32	c	18.2	38.7	6028	3	US-09-011-745-5	Sequence 5, Appli
33		18	38.3	495	1	US-08-406-248-1	Sequence 1, Appli
34		18	38.3	1194	4	PCT-US94-09700-10	Sequence 10, Appl
35		18	38.3	1335	2	US-08-985-090-3	Sequence 3, Appli
36		18	38.3	1335	3	US-09-165-543-3	Sequence 3, Appli
37		18	38.3	1335	3	US-09-167-354-6	Sequence 6, Appli
38	c	18	38.3	1721	1	US-08-241-766-3	Sequence 3, Appli
39		18	38.3	2100	1	US-08-154-915-5	Sequence 5, Appli
40		18	38.3	2106	1	US-07-970-462A-1	Sequence 1, Appli
41		18	38.3	2106	4	PCT-US92-10904-1	Sequence 1, Appli
42		18	38.3	2106	4	PCT-US94-09700-1	Sequence 1, Appli
43		18	38.3	2121	1	US-08-574-043A-1	Sequence 1, Appli
44		18	38.3	2121	2	US-08-795-015-1	Sequence 1, Appli
45		18	38.3	2121	4	PCT-US94-12936-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-136-251-1
; Sequence 1, Application US/09136251A
; Patent No. 6127156
; GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: MIYAZAKI, Taro
; APPLICANT: OJIMA, Setsuko
; APPLICANT: SHINJOH, Masako
; APPLICANT: TOMIYAMA, No. 6127156ibumi
; TITLE OF INVENTION: D-SORBITOL DEHYDROGENASE GENE
; FILE REFERENCE: D-Sorbital Dehydrogenase Gene
; CURRENT APPLICATION NUMBER: US/09/136,251A
; CURRENT FILING DATE: 1998-08-19
; EARLIER APPLICATION NUMBER: EP 97114432.4
; EARLIER FILING DATE: 1997-08-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3481
; TYPE: DNA
; ORGANISM: Gluconobacter suboxydans
US-09-136-251-1

Query Match 45.1%; Score 21.2; DB 3; Length 3481;
Best Local Similarity 76.5%; Pred. No. 10;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy      13 gccgcatattcggcctgatcggccgcaagcttgg 46
           || || ||||   || || ||||||| ||||||
Db     355 gcggcgtattcatgctcatgggccgcacgcttgg 388
```

RESULT 2
US-08-200-900A-26
; Sequence 26, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/200,900A
;       FILING DATE:  23-FEB-1994
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meinert, Maureen C.
;       REGISTRATION NUMBER:  31,544
;       REFERENCE/DOCKET NUMBER:  GI 5201-FWC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 876-1170 X8574
;       TELEFAX:  (617) 876-5851
;   INFORMATION FOR SEQ ID NO:  26:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  41 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-200-900A-26


  Query Match             43.4%;  Score 20.4;  DB 1;  Length 41;
  Best Local Similarity   71.1%;  Pred. No. 11;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       5 ataatgcggccgcatattcggcctgatcggccgcaagc 42
           | ||||||||||||  ||| ||||| |      |||||
Db       4 AGAATGCGGCCGCAAGTTCAGCCTGGTTAAGTCCAAGC 41


RESULT   3
PCT-US94-00616-26
; Sequence 26, Application PC/TUS9400616
;  GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  CLONING OF ENTEROKINASE AND METHOD OF USE
;     NUMBER OF SEQUENCES:  33
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US94/00616
;       FILING DATE:
;       CLASSIFICATION:
;   INFORMATION FOR SEQ ID NO:  26:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  41 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
PCT-US94-00616-26


  Query Match             43.4%;  Score 20.4;  DB 4;  Length 41;
  Best Local Similarity   71.1%;  Pred. No. 11;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       5 ataatgcggccgcatattcggcctgatcggccgcaagc 42
           | ||||||||||||  ||| ||||| |      |||||
Db       4 AGAATGCGGCCGCAAGTTCAGCCTGGTTAAGTCCAAGC 41


RESULT   4
US-08-948-705-2/c
; Sequence 2, Application US/08948705A
; Patent No. 6043084
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Matthew J.
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Old, Lloyd J.
;  TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
;  TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND METHODS FOR DIAGNOSING AND
;  TITLE OF INVENTION: TREATING COLON CANCER
;  FILE REFERENCE: LUD-5506-JEL/NDH
;  CURRENT APPLICATION NUMBER: US/08/948,705A
;  CURRENT FILING DATE: 1997-10-10
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2885
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
US-08-948-705-2


  Query Match             43.0%;  Score 20.2;  DB 3;  Length 2885;
  Best Local Similarity   75.8%;  Pred. No. 24;
  Matches   25;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      10 gcggccgcatattcggcctgatcggccgcaagc 42
           || ||||||  || |||||| |||||| | | |
Db    2747 GCTGCCGCAGCTTGGGCCTGTTCGGCCCCCACC 2715


RESULT   5
US-08-054-077C-1
; Sequence 1, Application US/08054077C
; Patent No. 5527679
;  GENERAL INFORMATION:
;     APPLICANT:  HEMLER, MARTIN E.
;     APPLICANT:  RAMASWAMY, HEMAVATHI
;     TITLE OF INVENTION:  HUMAN INTEGRIN  5 SUBUNIT PROTEIN
;     NUMBER OF SEQUENCES:  7
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;       ADDRESSEE:  CUSHMAN
;       STREET:  130 WATER STREET
;       CITY:  BOSTON
;       STATE:  MASSACHUSETTS
;       COUNTRY:  US
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/054,077C
;       FILING DATE:  27-APR-1993
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 07/694314
;       FILING DATE:  01-MAY-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  RESNICK, DAVID S.
;       REGISTRATION NUMBER:  34235
;       REFERENCE/DOCKET NUMBER:  40937
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 523-3400
;       TELEFAX:  (617) 523-6440
;       TELEX:  200291 STRE UR
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3415 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     FEATURE:
;       NAME/KEY:  mat peptide
```

```
;      LOCATION:  406..2733
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  337..2733
US-08-054-077C-1


  Query Match             42.1%;  Score 19.8;  DB 1;  Length 3415;
  Best Local Similarity   69.2%;  Pred. No. 35;
  Matches   27;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      8 atgcggccgcatattcggcctgatcggccgcaagcttgg 46
          |||| |  ||| |   |||| |||| || || ||| |||
Db   2115 ATGCCGGGGCAGAGATGGCCAGATCTGCAGCGAGCGTGG 2153


RESULT   6
US-08-629-643A-5
; Sequence 5, Application US/08629643A
; Patent No. 6025539
;  GENERAL INFORMATION:
;    APPLICANT:  Lee, J. L.
;    APPLICANT:  Lee, N. A.
;    TITLE OF INVENTION:  IL-5 TRANSGENIC MOUSE
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Schwegman, Lundberg, Woessner & Kluth, P. A.
;      STREET:  P. O. Box 2938
;      CITY:  Minneapolis
;      STATE:  MN
;      COUNTRY:  USA
;      ZIP:  55402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/629,643A
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/629,643
;      FILING DATE:  09-APR-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Viksnins, Ann S
;      REGISTRATION NUMBER:  37,748
;      REFERENCE/DOCKET NUMBER:  150.167WO1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  612-359-3260
;      TELEFAX:  612-359-3263
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6727 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  650...3771
;      OTHER INFORMATION:  Join 650..730, 1560..1592,
;      OTHER INFORMATION:  3468..3596, 3676..3771
US-08-629-643A-5


  Query Match             42.1%;  Score 19.8;  DB 3;  Length 6727;
  Best Local Similarity   63.8%;  Pred. No. 39;
  Matches   30;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy      1 agctataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
          ||| ||  | | | | | | |||  |||| || ||| || |  ||||
```

```
Db   5688 AGCGATCCTCCTGTCTCCTCTTCCTCCTGCTCAGCCCCAGGGGTGGA 5734


RESULT   7
US-09-280-799-1
; Sequence 1, Application US/09280799
; Patent No. 6136603
; GENERAL INFORMATION:
;  APPLICANT: Dean, Nicholas M.
;  APPLICANT: Karras, James G
;  APPLICANT: McKay, Robert
;  TITLE OF INVENTION: ANTISENSE MODULATION OF INTERLEUKIN-5 SIGNAL
;  TITLE OF INVENTION: TRANSDUCTION
;  FILE REFERENCE: ISPH-0340
;  CURRENT APPLICATION NUMBER: US/09/280,799
;  CURRENT FILING DATE: 1999-03-26
;  NUMBER OF SEQ ID NOS: 208
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 6727
;   TYPE: DNA
;   ORGANISM: Mus musculus
US-09-280-799-1


  Query Match             42.1%;  Score 19.8;  DB 3;  Length 6727;
  Best Local Similarity   63.8%;  Pred. No. 39;
  Matches   30;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy      1 agctataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
          ||| ||  | | | | | | |||  |||| || ||| || |  ||||
Db   5688 agcgatcctcctgtctcctcttcctcctgctcagccccaggggtgga 5734


RESULT   8
US-08-042-747A-7/c
; Sequence 7, Application US/08042747A
; Patent No. 5487969
;  GENERAL INFORMATION:
;    APPLICANT:  Eberle, Richard
;    APPLICANT:  Black, Darla
;    APPLICANT:  Scinicariello, Franco
;    APPLICANT:  Hilliard, Julia K.
;    TITLE OF INVENTION:  Cloning and Amplification of Monkey B
;    TITLE OF INVENTION:  Virus Genes
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cox & Smith Incorporated
;      STREET:  112 East Pecan Street, Suite 2000
;      CITY:  San Antonio
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  78205
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/042,747A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Haymond, W. Bradley
;      REGISTRATION NUMBER:  35186
;      REFERENCE/DOCKET NUMBER:  S-0072.179
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  210-554-5500
;      TELEFAX:  210-226-8395
;      TELEX:  767609
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2943 base pairs
```

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  Genomic DNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  87..2744
US-08-042-747A-7


  Query Match             41.3%;  Score 19.4;  DB 1;  Length 2943;
  Best Local Similarity   64.4%;  Pred. No. 49;
  Matches   29;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy       2 gctataatgcggccgcatattcggcctgatcggccgcaagcttgg 46
           || |  | |||| ||   | || | |    ||||||| |||||||
Db      61 GCGACTACGCGGGCGTCGACTCCGGCCCGCCGGCCGCGAGCTTGG 17


RESULT   9
US-08-469-667-1/c
; Sequence 1, Application US/08469667
; Patent No. 5733748
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Guo-Liang
;    APPLICANT:  Rosen, Craig
;    TITLE OF INVENTION:  Colon Specific Genes and Proteins
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi,
;      ADDRESSEE:  Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07068-1739
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,667
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferraro, Gregory D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  638 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..501
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..501
US-08-469-667-1


  Query Match             40.4%;  Score 19;  DB 1;  Length 638;
  Best Local Similarity   75.9%;  Pred. No. 56;
  Matches   22;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      15 cgcatattcggcctgatcggccgcaagct 43
```

```
             | ||||  ||||||| |:||| || ||||
Db      36 CACATACACGGCCTGGTSGGCAGCCAGCT 8


RESULT  10
PCT-US95-07289-1/c
; Sequence 1, Application PC/TUS9507289
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Guo-Liang
;    APPLICANT:  Rosen, Craig
;    TITLE OF INVENTION:  Colon Specific Genes and Proteins
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi,
;      ADDRESSEE:  Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07068-1739
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07289
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferraro, Gregory D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-265
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  638 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..501
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..501
PCT-US95-07289-1


  Query Match             40.4%;  Score 19;  DB 4;  Length 638;
  Best Local Similarity   75.9%;  Pred. No. 56;
  Matches   22;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      15 cgcatattcggcctgatcggccgcaagct 43
           | ||||  ||||||| |:||| || ||||
Db      36 CACATACACGGCCTGGTSGGCAGCCAGCT 8


RESULT  11
US-08-853-659A-28/c
; Sequence 28, Application US/08853659A
; Patent No. 5925522
;  GENERAL INFORMATION:
;    APPLICANT:  Wong, K.K.; Saffer, J.D.
;    TITLE OF INVENTION:  A Salmonella Sequence, Methods Of Detection
;    TITLE OF INVENTION:  Of A
;    TITLE OF INVENTION:  Salmonella Sequence, And Methods Of Detection Of Salmonella
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Paul W. Zimmerman
```

```
;      ADDRESSEE:  Intellectual Property Services
;      ADDRESSEE:  Battelle Memorial Institute
;      ADDRESSEE:  PNNL P.O. Box 999
;      STREET:  Washington Way
;      CITY:  Richland
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  99352
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette 3.50", 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Processor (WordPerfect 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/853,659A
;      FILING DATE:  Unknown
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  none
;      FILING DATE:  n/a
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1797 bases
;      TYPE:  nucleotide
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:  the coding nucleotides of SEQ ID
;      OTHER INFORMATION:  NO:28 correspond to nucleotides 22630 through
;      OTHER INFORMATION:  24426 of SEQ ID NO:2
US-08-853-659A-28


  Query Match             40.0%;  Score 18.8;  DB 2;  Length 1797;
  Best Local Similarity   68.4%;  Pred. No. 78;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      3 ctataatgcggccgcatattcggcctgatcggccgcaa 40
            || |||||| |  |  | ||| ||||||||  |  |||
Db   1101 CTGTAATGCCGGTGTTTTTTCTGCCTGATCTTCATCAA 1064


RESULT  12
US-08-853-659A-6/c
; Sequence 6, Application US/08853659A
; Patent No. 5925522
;  GENERAL INFORMATION:
;    APPLICANT:  Wong, K.K.; Saffer, J.D.
;    TITLE OF INVENTION:  A Salmonella Sequence, Methods Of Detection
;    TITLE OF INVENTION:  Of A
;    TITLE OF INVENTION:  Salmonella Sequence, And Methods Of Detection Of Salmonella
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Paul W. Zimmerman
;      ADDRESSEE:  Intellectual Property Services
;      ADDRESSEE:  Battelle Memorial Institute
;      ADDRESSEE:  PNNL P.O. Box 999
;      STREET:  Washington Way
;      CITY:  Richland
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  99352
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette 3.50", 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Processor (WordPerfect 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/853,659A
;      FILING DATE:  Unknown
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  none
;      FILING DATE:  n/a
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8967 bases
;      TYPE:  nucleotide
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
;    FEATURE:
;      OTHER INFORMATION:  SEQ ID NO:6 corresponds to
;      OTHER INFORMATION:  nucleotides 15735 through 24701 of SEQ ID NO:2
US-08-853-659A-6


  Query Match             40.0%;  Score 18.8;  DB 2;  Length 8967;
  Best Local Similarity   68.4%;  Pred. No. 98;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      3 ctataatgcggccgcatattcggcctgatcggccgcaa 40
            || |||||| |  |  | ||| ||||||||  |  |||
Db   7996 CTGTAATGCCGGTGTTTTTTCTGCCTGATCTTCATCAA 7959


RESULT  13
US-08-853-659A-9
; Sequence 9, Application US/08853659A
; Patent No. 5925522
;  GENERAL INFORMATION:
;    APPLICANT:  Wong, K.K.; Saffer, J.D.
;    TITLE OF INVENTION:  A Salmonella Sequence, Methods Of Detection
;    TITLE OF INVENTION:  Of A
;    TITLE OF INVENTION:  Salmonella Sequence, And Methods Of Detection Of Salmonella
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Paul W. Zimmerman
;      ADDRESSEE:  Intellectual Property Services
;      ADDRESSEE:  Battelle Memorial Institute
;      ADDRESSEE:  PNNL P.O. Box 999
;      STREET:  Washington Way
;      CITY:  Richland
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  99352
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette 3.50", 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Processor (WordPerfect 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/853,659A
;      FILING DATE:  Unknown
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  none
;      FILING DATE:  n/a
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8967 bases
;      TYPE:  nucleotide
;      STRANDEDNESS:  double stranded
;      TOPOLOGY:  linear
US-08-853-659A-9


  Query Match             40.0%;  Score 18.8;  DB 2;  Length 8967;
  Best Local Similarity   68.4%;  Pred. No. 98;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      3 ctataatgcggccgcatattcggcctgatcggccgcaa 40
            || |||||| |  |  | ||| ||||||||  |  |||
Db    972 CTGTAATGCCGGTGTTTTTTCTGCCTGATCTTCATCAA 1009


RESULT  14
```

```
US-08-853-659A-64/c
; Sequence 64, Application US/08853659A
; Patent No. 5925522
;  GENERAL INFORMATION:
;    APPLICANT:  Wong, K.K.; Saffer, J.D.
;    TITLE OF INVENTION:  A Salmonella Sequence, Methods Of Detection
;    TITLE OF INVENTION:  Of A
;    TITLE OF INVENTION:  Salmonella Sequence, And Methods Of Detection Of Salmonella
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Paul W. Zimmerman
;      ADDRESSEE:  Intellectual Property Services
;      ADDRESSEE:  Battelle Memorial Institute
;      ADDRESSEE:  PNNL P.O. Box 999
;      STREET:  Washington Way
;      CITY:  Richland
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  99352
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette 3.50", 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Processor (WordPerfect 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/853,659A
;      FILING DATE:  Unknown
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  none
;      FILING DATE:  n/a
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8967 bases
;      TYPE:  nucleotide
;      STRANDEDNESS:  single stranded
;      TOPOLOGY:  linear
US-08-853-659A-64


  Query Match             40.0%;  Score 18.8;  DB 2;  Length 8967;
  Best Local Similarity   68.4%;  Pred. No. 98;
  Matches   26;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 ctataatgcggccgcatattcggcctgatcggccgcaa 40
           || |||||| |  |  | ||| ||||||||  |  |||
Db    7996 CTGTAATGCCGGTGTTTTTTCTGCCTGATCTTCATCAA 7959


RESULT  15
US-08-853-659A-67
; Sequence 67, Application US/08853659A
; Patent No. 5925522
;  GENERAL INFORMATION:
;    APPLICANT:  Wong, K.K.; Saffer, J.D.
;    TITLE OF INVENTION:  A Salmonella Sequence, Methods Of Detection
;    TITLE OF INVENTION:  Of A
;    TITLE OF INVENTION:  Salmonella Sequence, And Methods Of Detection Of Salmonella
;    NUMBER OF SEQUENCES:  67
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Paul W. Zimmerman
;      ADDRESSEE:  Intellectual Property Services
;      ADDRESSEE:  Battelle Memorial Institute
;      ADDRESSEE:  PNNL P.O. Box 999
;      STREET:  Washington Way
;      CITY:  Richland
;      STATE:  Washington
;      COUNTRY:  U.S.A.
;      ZIP:  99352
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette 3.50", 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  Word Processor (WordPerfect 5.1)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/853,659A
;      FILING DATE:  Unknown
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  none
;      FILING DATE:  n/a
;  INFORMATION FOR SEQ ID NO:  67:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8967 bases
;      TYPE:  nucleotide
;      STRANDEDNESS:  single stranded
;      TOPOLOGY:  linear
US-08-853-659A-67


  Query Match             40.0%;  Score 18.8;  DB 2;  Length 8967;
  Best Local Similarity   47.4%;  Pred. No. 98;
  Matches   18;  Conservative    8;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 ctataatgcggccgcatattcggcctgatcggccgcaa 40
           |: :||:|| |  |  : ::| |||:||:|  |  |||
Db     972 CUGUAAUGCCGGUGUUUUUUCUGCCUGAUCUUCAUCAA 1009


Search completed: March  4, 2001, 12:05:19
Job time: 5584 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 12:03:42 ; Search time 1021.9 Seconds
(without alignments)
235.379 Million cell updates/sec

Title: US-09-101-423B-8
Perfect score: 47
Sequence: 1 agctataatgcggccgcata.....tgatcggccgcaagcttgga 47

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pr1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vi:*
29: gb_ba3:*
30: gb_in1:*
31: gb_in2:*
32: gb_in3:*
33: gb_pl3:*
34: gb_pr4:*
35: em_ba1:*
36: em_ba2:*
37: em_htg1:*
38: em_htg2:*
39: em_htg3:*
40: em_htg4:*
41: em_htg5:*
42: em_htg6:*
43: em_htg7:*
44: em_htg8:*
45: em_htg9:*
46: em_htg10:*
47: em_hum3:*
48: em_hum4:*
49: em_hum5:*
50: em_hum6:*
51: gb_pr5:*
52: gb_pr6:*
53: gb_pr7:*
54: gb_htg1:*
55: gb_htg2:*
56: gb_htg3:*
57: gb_htg4:*
58: gb_htg5:*
59: gb_htg6:*
60: gb_htg7:*
61: gb_htg8:*
62: gb_htg9:*
63: gb_htg10:*
64: gb_htg11:*
65: gb_htg12:*
66: gb_htg13:*
67: gb_htg14:*
68: gb_htg15:*
69: gb_htg16:*
70: gb_htg17:*
71: gb_htg18:*
72: gb_htg19:*
73: gb_htg20:*
74: gb_htg21:*
75: gb_htg22:*
76: gb_htg23:*
77: gb_sts1:*
78: gb_sts2:*
79: gb_vi1:*
80: gb_vi2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	24.8	52.8	86896	2	RCU57682	U57682 Rhodobacter
c	2	23.4	49.8	25458	29	SCE22	AL355832 Streptomy
	3	23.4	49.8	177241	52	HS402G11	AL022328 Human DNA
	4	23	48.9	2042	11	MUSGATGABA	L32178 Mus cookii
	5	23	48.9	4054	11	RATGABAT	M59742 Rat GABA tr
	6	23	48.9	4074	11	MUSGABAX	M92378 Mus musculu
	7	23	48.9	4402	11	MUSCOOK	M97512 Mus cookii
c	8	23	48.9	10908	1	AE002294	AE002294 Chlamydia
c	9	22.6	48.1	2013	2	BACCOTJABC	L38014 Bacillus su
	10	22.6	48.1	17103	59	AC017738	AC017738 Drosophil
c	11	22.6	48.1	18622	1	AF012532	AF012532 Bacillus
	12	22.6	48.1	101962	55	AC009915	AC009915 Drosophil
c	13	22.6	48.1	213190	2	BSUB0004	Z99107 Bacillus su
	14	22.6	48.1	226159	30	AE003738	AE003738 Drosophil
c	15	22.2	47.2	1950	2	AVIGLNA	M57275 A.vinelandi
c	16	22.2	47.2	2565	1	AF030293	AF030293 Rathayiba
c	17	22.2	47.2	2565	2	RRAJ2069	AJ002069 Rathayiba
c	18	22.2	47.2	6254	32	DMU13637	U13637 Drosophila
c	19	22.2	47.2	14724	58	AC015103	AC015103 Drosophil
c	20	22.2	47.2	151673	62	AC022365	AC022365 Homo sapi
	21	22.2	47.2	218971	56	AC011697	AC011697 Drosophil

c	22	22.2	47.2	308373	30	AE003495	AE003495 Drosophil
c	23	22	46.8	11936	1	AE004219	AE004219 Vibrio ch
c	24	21.8	46.4	1503	81	E15110	E15110 gDNA encodi
c	25	21.8	46.4	7647	81	E15111	E15111 Whole seque
	26	21.8	46.4	11627	1	AE004582	AE004582 Pseudomon
c	27	21.8	46.4	300950	2	AP001516	AP001516 Bacillus
	28	21.6	46.0	4566	2	FRANIFX	L29299 Frankia aln
c	29	21.6	46.0	47811	59	AC017664	AC017664 Drosophil
	30	21.6	46.0	119526	54	AC007468	AC007468 Drosophil
c	31	21.6	46.0	188167	68	AC068550	AC068550 Homo sapi
	32	21.6	46.0	210047	56	AC010906	AC010906 Homo sapi
c	33	21.6	46.0	341319	30	AE003536	AE003536 Drosophil
	34	21.4	45.5	5291	2	MLUSTROA	M17788 M.luteus st
c	35	21.4	45.5	11489	1	AE002455	AE002455 Neisseria
	36	21.4	45.5	35437	29	SCF43A	AL096837 Streptomy
	37	21.4	45.5	42655	29	SC7H2	AL109732 Streptomy
c	38	21.2	45.1	611	3	OAINIGFII6	U00667 Ovis aries
	39	21.2	45.1	2557	29	XANAG4G	D38228 Xanthomonas
	40	21.2	45.1	3481	2	AX000291	AX000291 Sequence
	41	21.2	45.1	11874	79	AHU18243	U18243 Alcelaphine
	42	21.2	45.1	21160	1	AF036940	AF036940 Pseudomon
	43	21.2	45.1	130608	79	AF005370	AF005370 Alcelaphi
	44	21.2	45.1	151802	73	AL354867	AL354867 Homo sapi
c	45	21	44.7	735	11	AF232928	AF232928 Mus muscu

ALIGNMENTS

RESULT 1
RCU57682/c
LOCUS RCU57682 86896 bp DNA BCT 07-FEB-1997
DEFINITION Rhodobacter capsulatus cosmids 143-147, complete sequence.
ACCESSION U57682
VERSION U57682.1 GI:1613791
KEYWORDS HTG.
SOURCE Rhodobacter capsulatus.
ORGANISM Rhodobacter capsulatus
Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group; Rhodobacter.
REFERENCE 1 (bases 1 to 86896)
AUTHORS Kumar,V., Fonstein,M. and Haselkorn,R.
TITLE Bacterium genome sequence
JOURNAL Nature 381 (6584), 653-654 (1996)
MEDLINE 96260014
REFERENCE 2 (bases 1 to 86896)
AUTHORS Fonstein,M., Kumar,V. and Haselkorn,R.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-1996) M. Fonstein, MGCB, Univ of Chicago, 920 E. 58th, Chicago, IL 60637, USA
FEATURES Location/Qualifiers
source 1. .86896
/organism="Rhodobacter capsulatus"
/strain="SB1003"
/db_xref="taxon:1061"
/map="cosmids 143-147 of the complete encyclopedia"
BASE COUNT 15089 a 28056 c 28787 g 14963 t 1 others
ORIGIN

Query Match 52.8%; Score 24.8; DB 2; Length 86896;
Best Local Similarity 80.6%; Pred. No. 50;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy      11 cggccgcatattcggcctgatcggccgcaagcttgg 46
           ||||||| |||||||| || ||||||| | ||||
Db   76687 CGGCCGCGCGTTCGGCCTCATAGGCCGCACGGTTGG 76652
```

RESULT 2
SCE22/c
LOCUS SCE22 25458 bp DNA BCT 10-MAY-2000
DEFINITION Streptomyces coelicolor cosmid E22.
ACCESSION AL355832
VERSION AL355832.1 GI:7799503
KEYWORDS ABC transporter integral membrane protein; bar; deoR-family transcriptional regulator; fructose-specific permease; hrdD; integral membrane protein; lipoprotein; MarR-family transcriptional repressor; mechanosensitive channel; methylthioadenosine phosphorylase; Na+/H+ antiporter; phosphofructokinase; secreted penicillin acylase; secreted protein; tetR-family transcriptional regulator; transmembrane efflux protein.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 25458)
AUTHORS Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
REFERENCE 2 (bases 1 to 25458)
AUTHORS Saunders,D.C. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 25458)
AUTHORS Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL; http://www.sanger.ac.uk/Projects/S_coelicolor/)
CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.
Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Cosmid E22 lies between E87 and E8 on the AseI-E genomic restriction fragment.
FEATURES Location/Qualifiers
source 1. .25458
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid E22"
gene complement(1. .490)
/gene="SCE22.01c"
CDS complement(<1. .490)

```
                    /gene="SCE22.01c"
                    /note="SCE22.01c, possible secreted penicillin acylase
                    (fragment), len: >162 aa; similar to C-terminal region of
                    SW:PAC2_PSES3 (EMBL:M18278) Pseudomonas sp. penicillin
                    acylase II precursor (EC 3.5.1.11) AcyII, 773 aa; fasta
                    scores: opt: 291 z-score: 347.5 E(): 6.8e-12; 39.1%
                    identity in 110 aa overlap. Contains possible N-terminal
                    signal peptide sequence"
                    /codon_start=1
                    /transl_table=11
                    /product="putative secreted penicillin acylase (fragment)"
                    /protein_id="CAB90968.1"
                    /db_xref="GI:7799504"
                    /translation="MPPTTTASTGQQAGTSGRKKARKGRKGRLLVLVLVLALIGGLAY
                    GAYWSISTVRASFPQTKGSITLDGLSGPVDVKRDGYGIPQVYASTEEDLFMAQGYVQA
                    QDRFYEMDVRRHMTAGRLSEMFGKSQIDNDEFLRTLGWHRVAKKEYDEKLSDSTKKYL
                    QAY"
misc_feature        1. .107
                    /note="nominal overlap with Streptomyces coelicolor cosmid
                    SCE87"
gene                805. .2352
                    /gene="SCE22.02"
CDS                 805. .2352
                    /gene="SCE22.02"
                    /note="SCE22.02, possible Na+/H+ antiporter, len: 515 aa;
                    similar to TR:Q9XAJ9 (EMBL:AL079348) Streptomyces
                    coelicolor putative Na(+)/H(+) antiporter SC66T3.14c, 514
                    aa; fasta scores: opt: 1828 z-score: 1992.7 E(): 0; 55.1%
                    identity in 497 aa overlap. Contains Pfam match to entry
                    PF00999 Na_H_Exchanger, Sodium/hydrogen exchanger family
                    and possible hydrophobic membrane spanning regions"
                    /codon_start=1
                    /transl_table=11
                    /product="putative Na+/H+ antiporter"
                    /protein_id="CAB90969.1"
                    /db_xref="GI:7799505"
                    /translation="MRCARERERPLTVHHLNQLLLVCSLVLLIAVAAVRISSRSGLPS
                    LLVYLAIGVAMGQDGIGDIQFDNAELVQVIGYGALVVILAEGGLGTKWKEAKPALPAA
                    SALALGGVAVSVGVTAAGAHYLTGLEWRQALIVGAVVSSTDAAAVFSVLRRIPLPKRI
                    TGTLEAESGFNDAPVVILVVAFSTAGPIEHWYVLLGEIALELAIGAAIGLAVGWLGSW
                    GLKHVALPASGLYPIAVMSIAIAAYAAGAMAHGSGFLAVYLASMVMGNARLPHWPATR
                    GFADGLGWLAQIGMFVLLGLLVTPHELGDDILPALVIGLVLTMVARPLSVVLCLVPFR
                    VPWQEQALMSWAGLRGAVPIILATIPMVEGVAGSHRIFNIVFVLVVVYTLVQGPTLPW
                    LARKLRLGKGDEAADLGIESAPLERLRGHLLSVTIPEGSRMHGVEVNELRLPTGAAVT
                    LVVRDGTSFVPLPTTGLRRGDELLVVATDPVRDAAEARLRAVGHGGKLAGWLGTGGTE
                    GARGNRR"
misc_feature        856. .2007
                    /gene="SCE22.02"
                    /note="Pfam match to entry PF00999 Na_H_Exchanger,
                    Sodium/hydrogen exchanger family, score 73.90, E-value
                    3.4e-18"
gene                2601. .3866
                    /gene="SCE22.03"
CDS                 2601. .3866
                    /gene="SCE22.03"
                    /note="SCE22.03, possible ABC transporter integral
                    membrane protein, len: 421 aa; similar to TR:O50532
                    (EMBL:AL009204) Streptomyces coelicolor hypothetical 41.1
                    kD protein SC9B10.25c, 393 aa; fasta scores: opt: 721
                    z-score: 787.1 E(): 0; 34.7% identity in 363 aa overlap.
                    Contains match to Prosite entry PS00211 ABC transporters
                    family signature and possible hydrophobic membrane
                    spanning regions"
                    /codon_start=1
                    /transl_table=11
                    /product="putative ABC transporter integral membrane
                    protein"
                    /protein_id="CAB90970.1"
                    /db_xref="GI:7799506"
                    /translation="MASTVTRPGYGQLLRTRGAWTFLLPGFAARQPFAMLTLSIVLLV
                    QHTTGSYGVAGAAAAVTGVSMAVFAPYSGRLADRYGQRAVLLPGVLVHTASGLTLTVL
                    ALADAPLWALFLAAVPTGASVPQVGPMVRARWAVQLKDSPLMSTAAAFESVTDELTFV
                    LGPLVATALCTAVDPAAGLVTEAALTLVGGLLFAARKSTEPKVGSADGGHARVEHVSA
                    LRVPGVRVLIVAFLGIGSVFGGMQVSLAAFTESIGEPGLNGVLYGVFAAGNMISGLAC
                    GAIAWKVAPQRRLLVGYAALALTASGLWAAHSVLVLAGLGLLVGMCVAPAIVTGYTLV
                    EDLVPAGARTEAFTWLTGAVALGQAAAVTVAGQLEDRFRDGAGFLVPMGGTVLALAVL
                    VALRSRLATRSHGRTVARGVGHRAPATVD"
misc_feature        3660. .3704
                    /gene="SCE22.03"
                    /note="PS00211 ABC transporters family signature"
RBS                 3923. .3928
gene                3934. .4266
                    /gene="SCE22.04"
CDS                 3934. .4266
                    /gene="SCE22.04"
                    /note="SCE22.04, hypothetical protein, len: 110 aa;
                    similar to TR:O05574 (EMBL:Z94752) Mycobacterium
                    tuberculosis hypothetical 11.4 kD protein MTCI237.05c, 110
                    aa; fasta scores: opt: 379 z-score: 372.3 E(): 2.8e-13;
                    52.3% identity in 107 aa overlap. High content in serine
                    amino acid residue in C-terminal domain"
                    /codon_start=1
                    /transl_table=11
                    /product="hypothetical protein SCE22.04"
                    /protein_id="CAB90971.1"
                    /db_xref="GI:7799507"
                    /translation="MPTYQYQCTECGEGLEAVQKFTDDALTECPNCQGRLKKVFSAVG
                    IVFKGSGFYRNDSRGSSSSSSPASSSSKTASSSSSDSGSSSSSSASSSGSGSSSGSGS
                    SSAGTTAA"
misc_feature        4111. .4251
                    /gene="SCE22.04"
                    /note="serine amino acid residues rich region"
gene                4397. .5239
                    /gene="SCE22.05"
CDS                 4397. .5239
                    /gene="SCE22.05"
                    /note="SCE22.05, possible methylthioadenosine
                    phosphorylase, len: 280 aa; similar to TR:CAB49985
                    (EMBL:AJ248286) Pyrococcus abyssi methylthioadenosine
                    phosphorylase related (MtaP), 265 aa; fasta scores: opt:
                    711 z-score: 861.9 E(): 0; 43.4% identity in 265 aa
                    overlap and to SW:MTAP_HUMAN (EMBL:U22233) Homo sapiens
                    5'-methylthioadenosine phosphorylase (EC 2.4.2.28)
                    (MTAPase) MtaP, 283 aa; fasta scores: opt: 653 z-score:
                    791.7 E(): 0; 38.2% identity in 254 aa overlap. Contains
                    Pfam match to entry PF00896 Mtap_PNP, phosphorylases
                    family 2"
                    /codon_start=1
                    /transl_table=11
                    /product="putative methylthioadenosine phosphorylase"
                    /protein_id="CAB90972.1"
                    /db_xref="GI:7799508"
                    /translation="MANKVKAEIGVIGGSGFYSFLDDVTEVRVDTPYGPPSDSLFLGE
                    VAGRRVAFLPRHGRGHHLPPHRINYRANLWALRSVGARQVLGPCAVGGLRPEYGPGTL
                    LVPDQFVDRTRSRPSTYFDGLPMPDGTVPNVVHVSLADPYCPTGRAAALKAARGREWE
```

```
  Query Match             49.8%;  Score 23.4;  DB 29;  Length 25458;
  Best Local Similarity   73.2%;  Pred. No. 1.9e+02;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      2 gctataatgcggccgcatattcggcctgatcggccgcaagc 42
          ||| |   ||||||| |||  || ||||   ||||||||||
Db  24364 GCTGTTCGGCGGCCGGATAGCCGACCTGTGGGGCCGCAAGC 24324


RESULT   3
HS402G11
LOCUS       HS402G11   177241 bp    DNA             PRI       12-DEC-1999
DEFINITION  Human DNA sequence from clone RP3-402G11 on chromosome
            22q13.31-13.33 Contains the MAPK12 gene for mitogen activated
            protein kinase 12 (SAPK3), the MAPK11 gene for mitogen activated
            protein kinase 11 (PRKM11), gene KIAA0315, the gene for a novel
            protein similar to KIAA0901 and mouse histone deacetylase MHDA2,
            the gene for a novel protein similar to Xenopus gamma-tubulin
            interacting protein (yeast SPC98 homolog), the gene for a novel
            protein similar to yeast and bacterial predicted proteins, the gene
```

```
            for a novel protein similar to C. elegans F38A5.2, the gene for a
            novel protein similar to MRS1 and the gene for a novel protein
            similar to mouse MOV10 (GB110) and yeast and plant predicted
            proteins. Contains ESTs, GSSs and fifteen putative CpG islands,
            complete sequence.
ACCESSION   AL022328
VERSION     AL022328.21  GI:5263010
KEYWORDS    HTG; CpG island; GB110; KIAA0315; KIAA0901; kinase; MAPK11; MAPK12;
            MHDA2; mitogen activated protein kinase; mitogen-activated; MOV10;
            MRS1; PRKM11; SAPK3; SPC98.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177241)
  AUTHORS   Coville,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 29, 1999 this sequence version replaced gi:5262835.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP3-402G11 is
            from the library RPCI-3 constructed at the Roswell Park Cancer
            Institute by the group of Pieter de Jong. For further details see
            http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP3-402G11 The true
            right end of clone RP5-898I4 is at 23403 in this sequence. The true
            right end of clone RP11-232E17 is at 56734 in this sequence. The
            true right end of clone RP4-600O24 is at 96013 in this sequence.
FEATURES                Location/Qualifiers
     source             1. .177241
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="22"
                        /map="q13.31-13.33"
                        /clone="RP3-402G11"
                        /clone_lib="RPCI-3"
     repeat_region      135. .447
                        /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region      657. .959
                        /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region      1373. .1659
                        /note="AluSx repeat: matches 1. .298 of consensus"
     repeat_region      1669. .1966
                        /note="AluSq repeat: matches 1. .306 of consensus"
     repeat_region      2504. .2574
                        /note="MER58 repeat: matches 170. .243 of consensus"
     repeat_region      2663. .2774
                        /note="L2 repeat: matches 2643. .2747 of consensus"
     repeat_region      2775. .3072
                        /note="AluSp repeat: matches 1. .300 of consensus"
     repeat_region      3073. .3086
                        /note="L2 repeat: matches 2630. .2643 of consensus"
     repeat_region      3959. .4455
                        /note="MER1A repeat: matches 31. .527 of consensus"
     repeat_region      5044. .5270
                        /note="AluJb repeat: matches 77. .303 of consensus"
     repeat_region      6305. .6592
                        /note="AluJo repeat: matches 29. .308 of consensus"
     misc_feature       6406. .6907
                        /note="match: GSS: Em:AQ475194"
     repeat_region      6853. .7155
                        /note="AluY repeat: matches 1. .303 of consensus"
     repeat_region      7218. .7516
                        /note="AluYb8 repeat: matches 1. .318 of consensus"
     repeat_region      9018. .9309
                        /note="AluSx repeat: matches 1. .294 of consensus"
     repeat_region      9427. .9542
                        /note="AluJo repeat: matches 30. .150 of consensus"
     repeat_region      9543. .9847
                        /note="AluY repeat: matches 1. .303 of consensus"
     repeat_region      9848. .10020
                        /note="AluJo repeat: matches 150. .306 of consensus"
     repeat_region      10024. .10316
                        /note="AluSg repeat: matches 1. .293 of consensus"
     repeat_region      10324. .10619
                        /note="AluY repeat: matches 2. .297 of consensus"
     repeat_region      10804. .11101
                        /note="AluSg repeat: matches 1. .296 of consensus"
     repeat_region      11227. .11641
                        /note="MER4C repeat: matches 22. .461 of consensus"
     repeat_region      11642. .11957
                        /note="AluY repeat: matches 1. .310 of consensus"
     repeat_region      11958. .11981
                        /note="MER4C repeat: matches 1. .22 of consensus"
     repeat_region      14063. .14370
                        /note="AluY repeat: matches 2. .310 of consensus"
     repeat_region      15488. .15718
                        /note="AluSx repeat: matches 51. .295 of consensus"
     repeat_region      16411. .16616
                        /note="AluY repeat: matches 91. .296 of consensus"
     repeat_region      16816. .17110
                        /note="AluJb repeat: matches 1. .305 of consensus"
     repeat_region      18096. .18300
                        /note="MER46C repeat: matches 113. .337 of consensus"
     repeat_region      18366. .18558
                        /note="MIR repeat: matches 32. .247 of consensus"
     repeat_region      18590. .18727
                        /note="LTR45 repeat: matches 389. .525 of consensus"
     repeat_region      21328. .21889
                        /note="MER54B repeat: matches 1. .638 of consensus"
     repeat_region      21958. .22115
                        /note="L1 repeat: matches 4613. .4770 of consensus"
     repeat_region      22116. .22245
                        /note="FLAM_C repeat: matches 1. .129 of consensus"
     repeat_region      22246. .22887
                        /note="L1 repeat: matches 3975. .4613 of consensus"
     repeat_region      22938. .23151
                        /note="FRAM repeat: matches 1. .166 of consensus"
     repeat_region      23152. .23217
                        /note="L1 repeat: matches 3919. .3994 of consensus"
     repeat_region      23218. .23373
                        /note="AluSg/x repeat: matches 133. .302 of consensus"
     repeat_region      23375. .23514
                        /note="L1 repeat: matches 3780. .3926 of consensus"
     repeat_region      23516. .23691
                        /note="FRAM repeat: matches -7. .168 of consensus"
     repeat_region      23699. .24482
                        /note="L1 repeat: matches 2915. .3769 of consensus"
     repeat_region      24501. .24747
                        /note="L1MEc repeat: matches 1501. .1751 of consensus"
     repeat_region      24978. .25194
                        /note="L1MEc repeat: matches 272. .492 of consensus"
     repeat_region      25933. .26246
                        /note="AluY repeat: matches 1. .309 of consensus"
```

```
repeat_region   27926. .28225
                /note="AluSx repeat: matches 2. .302 of consensus"
repeat_region   28694. .28997
                /note="AluSg repeat: matches 1. .304 of consensus"
repeat_region   29001. .29119
                /note="AluJo/FLAM repeat: matches 3. .133 of consensus"
repeat_region   29859. .30170
                /note="L1ME repeat: matches 5189. .5489 of consensus"
repeat_region   30171. .30449
                /note="AluSx repeat: matches 24. .304 of consensus"
repeat_region   30450. .30556
                /note="L1ME repeat: matches 5086. .5189 of consensus"
repeat_region   30567. .30813
                /note="L2 repeat: matches 1079. .1303 of consensus"
repeat_region   30975. .31047
                /note="L1MC/D repeat: matches 5325. .5397 of consensus"
repeat_region   31048. .31362
                /note="AluY repeat: matches 1. .303 of consensus"
repeat_region   31363. .31990
                /note="L1MC/D repeat: matches 4762. .5325 of consensus"
repeat_region   31997. .32351
                /note="MLT1A1 repeat: matches 16. .365 of consensus"
repeat_region   32365. .32907
                /note="L1MC/D repeat: matches 4184. .4755 of consensus"
repeat_region   32908. .33208
                /note="AluSx repeat: matches 1. .301 of consensus"
repeat_region   33209. .33464
                /note="L1MC/D repeat: matches 3936. .4184 of consensus"
repeat_region   33468. .33637
                /note="FRAM repeat: matches 1. .175 of consensus"
repeat_region   33650. .33707
                /note="L1MEc repeat: matches 2384. .2075 of consensus"
repeat_region   33708. .33998
                /note="AluSx repeat: matches 1. .288 of consensus"
repeat_region   33999. .34144
                /note="L1MEc repeat: matches 2231. .2385 of consensus"
repeat_region   34781. .34916
                /note="4 copies 34 mer 86 conserved"
repeat_region   34792. .34910
                /note="7 copies 17 mer 73 conserved"
repeat_region   35157. .35466
                /note="AluSg repeat: matches 1. .310 of consensus"
repeat_region   35626. .35922
                /note="AluSg repeat: matches 1. .297 of consensus"
repeat_region   35932. .36042

  Query Match             49.8%;  Score 23.4;  DB 52;  Length 177241;
  Best Local Similarity   73.2%;  Pred. No. 1.4e+02;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      6 taatgcggccgcatattcggcctgatcggccgcaagcttgg 46
          || ||||||| |  |||||||| | ||| | |  |||| ||
Db 101609 TACTGCGGCCACCAATTCGGCCAGTTCGCCGGGCAGCTGGG 101649


RESULT   4
MUSGATGABA
LOCUS       MUSGATGABA    2042 bp    mRNA            ROD       09-SEP-1994
DEFINITION  Mus cookii GABA transporter (GAT-1) mRNA, complete cds.
ACCESSION   L32178
VERSION     L32178.1  GI:533225
KEYWORDS    GABA transporter; gamma-aminobutyric acid transporter.
SOURCE      Mus cookii (library: lambda ZAP) adult retina cDNA to mRNA.
  ORGANISM  Mus cookii
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2042)
  AUTHORS   Ruiz,M., Egal,H., Qian,X.J., Sarthy,V.P. and Sarkar,H.K.
  TITLE     Cloning, expression and localization of a mouse retinal
            gamma-aminobutyric acid transporter
  JOURNAL   Unpublished (1994)
FEATURES             Location/Qualifiers
     source          1. .2042
                     /organism="Mus cookii"
                     /db_xref="taxon:10098"
                     /dev_stage="adult"
                     /tissue_type="retina"
                     /tissue_lib="lambda ZAP"
     gene            93. .1889
                     /gene="GAT-1"
     CDS             93. .1889
                     /gene="GAT-1"
                     /note="gamma-aminobutyric acid transporter"
                     /codon_start=1
                     /product="GABA transporter"
                     /protein_id="AAA37663.1"
                     /db_xref="GI:533226"
                     /translation="MATDNSKVADGQISTEVSEAPVASDKPKTLVVKVQKKAGDLPDR
                     DTWKGRFDFLMSCVGYAIGLGNVWRFPYLCGKNGGGAFLIPYFLTLIFAGVPLFLLEC
                     SLGQYTSIGGLGVWNVAPMFKGVGLAAAVLSFWLNIYYIVIISWAIYYLYNSFTTTLP
                     WKQCDNPWNTDRCFSNYSLVNTTNMTSAVVEFWERNMHQMTDGLDKPGQIRCLAITLA
                     IAWVLVYFCIWKGVGWTGKVVYFSATYPYIMLIILFFRGVTLPGAKEGILFYITPNFR
                     KLSDSEVWLDAATQIFFSYGLGLGSLIALGSYNSFHNNVYRDSIIVCCINSCTSMFAG
                     FVIFSIVGFMAHVTKRSIADVAASGPGLAFLAYPEAVTQLPISPLWAILFFSMLLMLG
                     IDSQFCTVEGFITALVDEYSRLLRNRRELFIAAVCIVSYLIGLSNITQGGIYVFKLFD
                     YYSASGMSLLFLVFFECVSISWFYGVNRFYDNIQEMVGSRPCIWWKLCWSFFTPIIVA
                     GVFLFSAVQMTPLTMGSYVFPKWGQGVGWLMALSSMVLIPGYMAYMFLTLKGSLKQRL
                     QVMIQPSEDIVRPENGPEQPQAGSSASKEAYI"
BASE COUNT      384 a    614 c    543 g    501 t
ORIGIN


  Query Match             48.9%;  Score 23;  DB 11;  Length 2042;
  Best Local Similarity   68.1%;  Pred. No. 3.6e+02;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 agctataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
          || |||  |||| ||  | | | | ||||| | |||||| ||| | |
Db   1814 AGATATTGTGCGCCCTGAGAATGGCCCTGAGCAGCCGCAGGCTGGCA 1860


RESULT   5
RATGABAT
LOCUS       RATGABAT     4054 bp    mRNA            ROD       07-MAR-1995
DEFINITION  Rat GABA transporter protein mRNA, complete cds.
ACCESSION   M59742 M33003
VERSION     M59742.1  GI:204221
KEYWORDS    GABA transporter; gamma-aminobutyric acid transporter.
SOURCE      Rat brain, cDNA to mRNA, clone GAT-1.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 4054)
  AUTHORS   Guastella,J., Nelson,N., Nelson,H., Czyzyk,L., Keynan,S.,
            Miedel,M.C., Davidson,N., Lester,H.A. and Kanner,B.I.
  TITLE     Cloning and expression of a rat brain GABA transporter
  JOURNAL   Science 249 (4974), 1303-1306 (1990)
  MEDLINE   90378307
FEATURES             Location/Qualifiers
     source          1. .4054
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="GAT-1"
                     /tissue_type="brain"
     gene            150. .4036
                     /gene="GABA transporter"
     CDS             150. .1949
                     /gene="GABA transporter"
                     /codon_start=1
                     /product="GABA transporter protein"
                     /protein_id="AAA63487.1"
                     /db_xref="GI:204222"
                     /translation="MATDNSKVADGQISTEVSEAPVASDKPKTLVVKVQKKAGDLPDR
                     DTWKGRFDFLMSCVGYAIGLGNVWRFPYLCGKNGGGAFLIPYFLTLIFAGVPLFLLEC
                     SLGQYTSIGGLGVWKLAPMFKGVGLAAAVLSFWLNIYYIVIISWAIYYLYNSFTTTLP
```

```
                     WKQCDNPWNTDRCFSNYSLVNTTNMTSAVVEFWERNMHQMTDGLDKPGQIRWPLAITL
                     AIAWVLVYFCIWKGVGWTGKVVYFSATYPYIMLIILFFRGVTLPGAKEGILFYITPNF
                     RKLSDSEVWLDAATQIFFSYGLGLGSLIALGSYNSFHNNVYRDSIIVCCINSCTSMFA
                     GFVIFSIVGFMAHVTKRSIADVAASGPGLAFLAYPEAVTQLPISPLWAILFFSMLLML
                     GIDSQFCTVEGFITALVDEYPRLLRNRRELFIAAVCIVSYLIGLSNITQGGIYVFKLF
                     DYYSASGMSLLFLVFFECVSISWFYGVNRFYDNIQEMVGSRPCIWWKLCWSFFTPIIV
                     AGVFLFSAVQMTPLTMGSYVFPKWGQGVGWLMALSSMVLIPGYMAYMFLTLKGSLKQR
                     LQVMIQPSEDIVRPENGPEQPQAGSSASKEAYI"
     TATA_signal     4031. .4036
                     /gene="GABA transporter"
BASE COUNT      885 a   1119 c    980 g   1070 t
ORIGIN


  Query Match             48.9%;  Score 23;  DB 11;  Length 4054;
  Best Local Similarity   68.1%;  Pred. No. 3.3e+02;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 agctataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
          || |||  |||| ||  | | | | ||||| | |||||| ||| | |
Db   1874 AGATATCGTGCGCCCTGAGAATGGTCCTGAGCAGCCGCAGGCTGGCA 1920


RESULT   6
MUSGABAX
LOCUS       MUSGABAX     4074 bp    mRNA            ROD       27-APR-1993
DEFINITION  Mus musculus GABA transporter mRNA sequence.
ACCESSION   M92378
VERSION     M92378.1  GI:193411
KEYWORDS    GABA transporter.
SOURCE      Mus musculus (library: cDAN lambda-zap) brain cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4074)
  AUTHORS   Liu,Z.-R., Mandiyan,S., Nelson,H. and Nelson,N.
  TITLE     A family of genes encoding neurotransmitter transporters
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89, 6639-6643 (1992)
  MEDLINE   92335351
FEATURES             Location/Qualifiers
     source          1. .4074
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="brain"
                     /tissue_lib="cDAN lambda-zap"
     misc_feature    1. .4074
                     /note="putative cds"
                     /product="GABA transporter"
BASE COUNT      887 a   1115 c    995 g   1077 t
ORIGIN


  Query Match             48.9%;  Score 23;  DB 11;  Length 4074;
  Best Local Similarity   68.1%;  Pred. No. 3.3e+02;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 agctataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
          || |||  |||| ||  | | | | ||||| | |||||| ||| | |
Db   1863 AGATATTGTGCGCCCTGAGAATGGCCCTGAGCAGCCGCAGGCTGGCA 1909


RESULT   7
MUSCOOK
LOCUS       MUSCOOK      4402 bp    DNA             ROD       27-APR-1993
DEFINITION  Mus cookii DNA sequence.
ACCESSION   M97512
VERSION     M97512.1  GI:192680
KEYWORDS    .
SOURCE      Mus cookii DNA.
  ORGANISM  Mus cookii
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4402)
  AUTHORS   Liu,Q.-R., Mandiyan,S., Nelson,H. and Nelson,N.
  TITLE     A family of genes encoding neurotransmitter transporters
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. (1992) In press
FEATURES             Location/Qualifiers
     source          1. .4402
                     /organism="Mus cookii"
                     /db_xref="taxon:10098"
BASE COUNT     1055 a   1090 c   1230 g   1027 t
ORIGIN


  Query Match             48.9%;  Score 23;  DB 11;  Length 4402;
  Best Local Similarity   68.1%;  Pred. No. 3.3e+02;
  Matches   32;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 agctataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
          || |||  |||| ||  | | | | ||||| | |||||| ||| | |
Db   3973 AGATATTGTGCGCCCTGAGAATGGCCCTGAGCAGCCGCAGGCTGGCA 4019


RESULT   8
AE002294/c
LOCUS       AE002294     10908 bp    DNA             BCT       26-MAY-2000
DEFINITION  Chlamydia muridarum, section 26 of 85 of the complete genome.
ACCESSION   AE002294 AE002160
VERSION     AE002294.1  GI:7190305
KEYWORDS    .
SOURCE      Chlamydia muridarum.
  ORGANISM  Chlamydia muridarum
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE   1  (bases 1 to 10908)
  AUTHORS   Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
            White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
            Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
            Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
            McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
  TITLE     Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
            pneumoniae AR39
  JOURNAL   Nucleic Acids Res. 28 (6), 1397-1406 (2000)
  MEDLINE   20150255
   PUBMED   10684935
REFERENCE   2  (bases 1 to 10908)
  AUTHORS   Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
            White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
            Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
            Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
            McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10908
                     /organism="Chlamydia muridarum"
                     /db_xref="taxon:83560"
                     /note="synonym: Chlamydia trachomatis MoPn"
     gene            241. .495
                     /gene="TC0265"
     CDS             241. .495
                     /gene="TC0265"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF39134.1"
                     /db_xref="GI:7190306"
                     /translation="MIIVGALRDPSWLYYFFLDPFTHVFEKSLSKTRNNKKGYSLFQI
                     FWGHQKSCFLFHKVKVSTWFVPYKVVGFFVQRLLDHLLKK"
     gene            426. .1016
                     /gene="TC0266"
     CDS             426. .1016
                     /gene="TC0266"
                     /note="hypothetical protein; identified by Glimmer2;
```

putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39135.1"
/db_xref="GI:7190307"
/translation="MVCPLQGCGLLRSEASGSSSEEVNQDPLLDAISYSVRVCLAVAK
HTESAKIYHQNRKRCHKVGCWGKVKRFLCCCCCAEKEWVEGKITGLGLWMQEHIRLHG
ILITGYAVHLSGVCWENMLLAGETLSPGDQRQLNAALENSSITWMRLMLDSSDGKVDG
LFSPAMQEDLFAQYREKVGIGNLADAVALRHFCNRT"

gene complement(1104. .3707)
/gene="TC0267"

CDS complement(1104. .3707)
/gene="TC0267"
/note="similar to GP:4376270; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF39136.1"
/db_xref="GI:7190308"
/translation="MRPDHVNLCCLCATILSPTAILFGQDALDKSALITKNPNSIVCT
FLEDCTMENFSPALLSHARQDDPLYIIGNTHNWFVSNLHPSTNEERFLKEKGDLSIQD
FRFLSFTDCSSSTEDSPSILYHKNGQLFLRNNGNMSFYRNHSEGSGGALSTDALFLQH
NYLFTNFEENSSAKNGGAIQAQTLSLSRNVSSLSFSRNRANLNGGAICCQNLICSGNV
NPLFFTNNSALNGGAICCINEQNLSEKGCLSLAYNQETLFSGNSAKEKGGAIYTKHMV
LRHNGPVSFVNNSAKLGGAIAIQSGGSLSIIAGGGSVLFQNNSCHFSDQGTVRNAIYL
EKNALLSSLEARHGDILFFDPIVQEVVSPEFSTTSALTPLRIQTNTNRAVIFSSENLS
KEEKTEANLISKIQQPIELQSGCLVLKDRVILSAPSLSQAPQALLVMDVGTSLTTSSD
LKLTTLSIPLHSIDTENSVSIQSPTLSIQKIFLSNSEHENFYENVELLSKDQKDIPLL
SLPKGLPHPDLPDGNLSSHFGYQGDWNFSWQTSDQRETLVANWTANSYIPHPERQSAL
VANTLWNTYSDMQAVQSMINTTAQGGAYLFGTWGSAVSNLFYSHGNSGKSTDNWKHRS
LGYLFGISTHSLDDHSFCLAAGQLFGKSSDSFVTSADTTSYIAAIQTQIATSLIKISA
QACYNESIHELKTKYRSFSKEGFGAWHSVAVSGEIGASIPIVSNGSGLFSSFSIFSKL
QGFSGKQDGFEESRGEARAFADSSFTNISLPVGIAFEKKSQKTRNYYHFLGAYIQDLK
RCVESGPVTLLKNSVTWDAPMANLDSRAWMFRLTNQRALHRFQTLVNMSYMLRGQSYS
YSLDLGTTYRF"

gene 3956. .5800
/gene="TC0268"

CDS 3956. .5800
/gene="TC0268"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39137.1"
/db_xref="GI:7190309"
/translation="MFYFLGWFVMGIKGVGGSGHSDYPIPSHNGDGESEKNSSDSTSS
KVNAKVTSSLQGAPSTNDENSVSPYSVVDVTDLIESGESSRHVIKKSIETEEAAHRES
SVEGAGHSSRGIFGRLQAGLGRLARRVGEAVRNTVGSIFPQRAGAEQRTGKARTKYSP
SASRGLRLMFTDFWRYRVLHRNPPMDGLFAKLDADEAEDMAAYTKEYVSNLEKRGAAD
RETIEHCQMVAKNWEKRARDLRDMGAAKKFLRDPFGKSDPKYKGTLPGEYTVGNTMFY
DGPGVSKLSEVDTGFWLDMEKLSDAVLSANIQKGLRARFVLNQSIPQLESLEERFRKL
ESACDEARASLKEAGWIKEGKEPNKAQRAFRRFVEESRNLELSFGSFGESARRLSARV
SQGLAAAGEAIRRCFDCRKGKYSLKKDLSSEELNLAEELIRFTDEMGIERDPDGNYNI
PWVENWRTGVPVIEGEGAEHIYETMMPVQESFEQVYEVMDMGLEERRDFAVSQQHYQV
PPRSSLNYETPRFREYDVPRNSARSYYDVPRVPPQNEVEEMHVTKGMRSSVYACFVAG
MRNYIVSQPQEQIPNSEQVEQLFQELINDGDQIIQELMKIWNEELDNQ"

gene complement(5923. .6192)
/gene="TC0269"

CDS complement(5923. .6192)
/gene="TC0269"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39138.1"
/db_xref="GI:7190310"
/translation="MLCKVCRGLSSLIVVLGAINAGILGATGYRVNIVARLLGEGTML
TQAAYVVAGIAGVIAFLNFFKCCFKKRHSDCCSSKGGCHHHTDRE"

gene 6393. .6695
/gene="TC0270"

CDS 6393. .6695
/gene="TC0270"
/note="glutamyl-tRNA(Gln) amidotransferase subunit C,
putative; identified by match to TIGR protein family HMM
TIGR00135"
/codon_start=1
/transl_table=11
/product="glutamyl-tRNA(Gln) amidotransferase subunit C,
putative"
/protein_id="AAF73542.1"
/db_xref="GI:8163182"
/translation="MTQSYVNKEEIISLAKNAALELEEAHVEEFVTSMNNVIALMQEV
VAIDISDIILEATVHHFVGPEDLREDMVTSDFTREEFLANVPVSLGGLVKVPTVIK"

gene 6707. .8182
/gene="TC0271"

CDS 6707. .8182
/gene="TC0271"
/note="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamyl-tRNA(Gln) amidotransferase subunit A"
/protein_id="AAF39139.1"
/db_xref="GI:7190311"
/translation="MYRKSALELRDAVVNREISVTAITEYFYHRIESYDEQIGSFLSL
CKERALLRASRIDDKLAKGDPIGILAGIPIGVKDNIHITGVKTTCASKMLENFVAPFD
ATVVRRIEMEDGILLGKLNMDEFAMGSTTRYSAFQHTNNPWDLERVPGGSSGGSAAAV
SARFCPIALGSDTGGSIRQPAAFCGVVGFKPSYGAVSRYGLVAFGSSLDQIGPLTTVV
EDVALAMDAFAGRDIKDATTRDFFRGTFSQALSLEVPKLIGVPRGFLDGLQEDCKENF
FEALAVMERQGSRIIDIDLSVLKHAVPVYYIVASAEAATNLARFDGVRYGHRCAQADN
MQEMYARSRKEGFGKEVTRRILLGNYVLSAERQNIFYKKGTAVRATLIEAFQSAFECC
DVIAMPVCASPAIRDTDVLDPVSLYLQDIYTVAVNLAYLPAISVPSGLSKEGLPLGVQ
FIGKRGADQQICQVGYSFQEHSQIKQLYPKAVNGLFDGGME"

gene 8184. .9650
/gene="TC0272"

CDS 8184. .9650
/gene="TC0272"
/note="similar to PID:1653395; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamyl-tRNA(Gln) amidotransferase subunit B"
/protein_id="AAF39140.1"
/db_xref="GI:7190312"
/translation="MGIAHTEWESVIGLEVHVELNTASKLFSPARNHFGDEPNTNISP
VCTGMPGALPVLNKDAVRKAVLFGCAVEGDVALFSRFDRKSYFYPDSPRNFQITQYEH
PIVRGGCVRAIVEGEEKTFELAQTHLEDDAGMLKHFGDFAGVDYNRAGVPLIEIVSKP
CMFSAEDAVAYANALVSILSYIGISDCNMEEGSVRFDVNISVRPKGSKELRNKVEIKN
MNSFTFMAQALEAEKHRQIEEYLSHPNEDPKKVVPAATYRWDPEKKKTVLMRLKERAE
DYMYFVEPDLPVLQITESYIDEVRQTLPELPHSKYMRYITDFDLAEDLAMILVSDRHT
AHFFETATMSCKNYRALSNWITVEFAGRCKATGKTLPFTGILPEWVAQLVNFIDRGVI
TGKIAKEIADKMVSSFGESPEDILRRHPSLLPMTDDHALRAIVKEVVAQNAASVEDYK
NGKAKALGFLVGQIMKRTEGKAPPKRVNELLLAAMRDA"

gene complement(9731. .10834)
/gene="TC0273"

CDS complement(9731. .10834)
/gene="TC0273"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39141.1"
/db_xref="GI:7190313"
/translation="MVKIMAPITPTTSPQVKGLLSRFLTAPDRHPKLRYVYDISLIAI
SILCIVSIILWTQGSGLALFAIAPALAIGALGVTLLVSDLAESPKSKEVADTVAAVSL
PFILTGTAAGLMFSAIAVGGGAVILANPLFLMGSMTLGFALMSLHKVTYQYLSNRSQW
QKQNKIKQIESAAWENKLPKESKESSLQTSVRYSSLARKDKTRRNKPGMPNKGSQVPA
SIANTERSLRSEEVLHSQSLLRQKELFPNTSNIKKELPNTKSILHTPLNRRSPSGSDS
DDVYYTPRAGLSSAETSALGDISGISSSSTSSKTSTPKAKRRVVRSSRSERNARHHRN
KEDHRQNQEESSDDEDSSPLPSPRRKKYRSRPK"

BASE COUNT 3210 a 1860 c 2587 g 3251 t

```
ORIGIN


  Query Match             48.9%;  Score 23;  DB 1;  Length 10908;
  Best Local Similarity   74.4%;  Pred. No. 2.9e+02;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       8 atgcggccgcatattcggcctgatcggccgcaagcttgg 46
           | || |||  || ||| |||| ||||||  ||||||| |
Db    4562 AAGCTGCCATATCTTCAGCCTCATCGGCATCAAGCTTTG 4524


RESULT   9
BACCOTJABC/c
LOCUS       BACCOTJABC   2013 bp    DNA             BCT       15-JUN-1996
DEFINITION  Bacillus subtilis cotJABC operon, polypeptide(s) affecting spore
            coat composition, cds.
ACCESSION   L38014
VERSION     L38014.1  GI:1377748
KEYWORDS    .
SOURCE      Bacillus subtilis (strain 168) DNA.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 2013)
  AUTHORS   Henriques,A.O., Beall,B.W., Roland,K. and Moran,C.P. Jr.
  TITLE     Characterization of cotJ, a sigma E-controlled operon affecting the
            polypeptide composition of the coat of Bacillus subtilis spores
  JOURNAL   J. Bacteriol. 177 (12), 3394-3406 (1995)
  MEDLINE   95286532
FEATURES             Location/Qualifiers
     source          1. .2013
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            326. .574
                     /gene="cotJA"
     CDS             326. .574
                     /gene="cotJA"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB17588.1"
                     /db_xref="GI:1377749"
                     /translation="MKDMQPFTPVKSYTPFHSRFDPCPPIGKKYYRTPPNLYMTFQPE
                     HMEQFSPMEALRKGTLWKDLYDFYENPYRGGDAHGKKG"
     gene            519. .821
                     /gene="cotJB"
     CDS             519. .821
                     /gene="cotJB"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB17589.1"
                     /db_xref="GI:1377750"
                     /translation="MIFMKTLIEGETHMAKKVDAEYYRQLEQIQAADFVLVELSLYLN
                     THPHDEDALKQFNQYSGYSRHLKRQFESSYGPLLQFGNSPAGKDWDWGKGPWPWQV"
     gene            836. .1405
                     /gene="cotJC"
     CDS             836. .1405
                     /gene="cotJC"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB17590.1"
                     /db_xref="GI:1377751"
                     /translation="MWVYEKKLQYPVKVSTCNPTLAKYLIEQYGGADGELAAALRYLN
                     QRYTIPDKVIGLLTDIGTEEFAHLEMIATMVYKLTKDATPEQLREAGLGDHYVNHDSA
                     LFYHNAAGVPFTASYIQAKGDPIADLYEDIAAEEKARATYQWLIDISDDPDLNDSLRF
                     LREREIVHSMRFREAVEILKEERDKKKIF"
BASE COUNT      515 a    426 c    553 g    519 t
ORIGIN
```

```
  Query Match             48.1%;  Score 22.6;  DB 2;  Length 2013;
  Best Local Similarity   68.9%;  Pred. No. 5e+02;
  Matches   31;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       3 ctataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
           | | |||| || | |||| | || ||| | ||||||| ||  ||
Db     190 CCAAAATGAGGTCCAATATGCAGCATGAGCAGCCGCAAACTCTGA 146


RESULT  10
AC017738
LOCUS       AC017738    17103 bp    DNA             HTG       10-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017738
VERSION     AC017738.1  GI:6554261
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 17103)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10211936 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .17103
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     5111 a   3705 c   3714 g   4573 t
ORIGIN


  Query Match             48.1%;  Score 22.6;  DB 59;  Length 17103;
  Best Local Similarity   68.9%;  Pred. No. 3.8e+02;
  Matches   31;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       1 agctataatgcggccgcatattcggcctgatcggccgcaagcttg 45
           ||| || | |||  ||| | |||||||| | || |  | ||| |||
Db   12160 AGCGATCAGGCGTTCGCCTGTTCGGCCAGCTCCGTGGAAAGTTTG 12204


RESULT  11
AF012532/c
LOCUS       AF012532    18622 bp    DNA             BCT       12-AUG-1997
DEFINITION  Bacillus subtilis strain 168 trpC2 YefA (yefA) gene, partial cds,
            and YefB (yefB), YefC (yefC), YeeA (yeeA), YeeB (yeeB), YeeC
            (yeeC), YeeD (yeeD), YeeE (yeeE) and YeeF (yeeF) genes, complete
            cds.
ACCESSION   AF012532
VERSION     AF012532.1  GI:2318058
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 15361 to 18622)
  AUTHORS   Henriques,A.O., Beall,B.W., Roland,K. and Moran,C.P. Jr.
  TITLE     Characterization of cotJ, a sigma E-controlled operon affecting the
            polypeptide composition of the coat of Bacillus subtilis spores
  JOURNAL   J. Bacteriol. 177 (12), 3394-3406 (1995)
  MEDLINE   95286532
REFERENCE   2  (bases 1 to 18622)
  AUTHORS   Borriss,R. and Schroeter,R.
```

```
  TITLE     The 55-58 degree segment of the Bacillus subtilis chromosome, a
            region spanning from the purA gene cluster to the cotJ operon
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 18622)
  AUTHORS   Borriss,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1997) Borriss R., Humboldt University, Institute
            of Biology, Chausseestrasse 117, Berlin, Germany, D-10115
FEATURES             Location/Qualifiers
     source          1. .18622
                     /organism="Bacillus subtilis"
                     /strain="168 trpC2"
                     /db_xref="taxon:1423"
                     /map="55 to 58 degree"
     gene            <1. .81
                     /gene="yefA"
     CDS             <1. .81
                     /gene="yefA"
                     /note="similar to E.coli hypothetical RNA
                     methyltransferase: SwissProt Accession Number P55135;
                     similar to Chlamydia trachomatis Hom1: SwissProt Accession
                     Number P55137"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YefA"
                     /protein_id="AAB66480.1"
                     /db_xref="GI:2318067"
                     /translation="AKKRLRRSPPADNCGDEAEAGSVCVL"
     gene            complement(227. .1140)
                     /gene="yefB"
     CDS             complement(227. .1129)
                     /gene="yefB"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YefB"
                     /protein_id="AAB66472.1"
                     /db_xref="GI:2318059"
                     /translation="MYNEGVTSPSQLARKKNATDKWHGSSIRKILENPHYTGTLQQCR
                     EYRPSVTSKRRRSVNLENQIIIEKSHEPIIPLEDFLIVQEILITRKRRRPQAEVHLFT
                     NTAVCKDCRRSMHFKKNRKGYVCGSYNKHGLKACSDHYVSEIDLTDKVITNLNKIYLK
                     FSKENYFKELSEKALKYKEKIESKIIEINEKLNDKKRDKSNLVISLANGVISKEDYQL
                     AINITNEDISNFETTLHQLSKDLEFQKIEKEIIDFKKSLDKFMKEGTLTPEMLHLLVD
                     EIEVHANGTIEINYRFREPTVPSA"
gene                 1024. .1445
                     /gene="yefC"
RBS                  1024. .1030
                     /gene="yefC"
CDS                  1068. .1445
                     /gene="yefC"
                     /note="similar to Enterococcus faecalis plasmid PY117 orf4
                     encoded by GenBank Accession Number D78257"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YefC"
                     /protein_id="AAB66473.1"
                     /db_xref="GI:2318060"
                     /translation="MSVAFFLRASCEGLVTPSLYNPLAIASKPFPAIYSEKILIFTIF
                     SAFAEFERDMIVERTQEGKMLAKQNPDFREGRPKKFTKQQINHALTLLENHSYKQVED
                     MTGISVSTLVRAKKKKAAEAING"
RBS                  complement(1133. .1140)
                     /gene="yefB"
gene                 1505. .4159
                     /gene="yeeA"
RBS                  1505. .1512
                     /gene="yeeA"
CDS                  1520. .4159
                     /gene="yeeA"
                     /note="similar to Methanococcus janaschii plasmid pURB801
                     hypothetical protein MJECS02 encoded by GenBank Accession
                     Number L77119"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YeeA"
                     /protein_id="AAB66474.1"
                     /db_xref="GI:2318061"
                     /translation="MALIDLEDKIAEIVNREDHSDFLYELLGVYDVPRATITRLKKGN
                     QNLTKRVGEVHLKNKVWFKEAKKGKLFDALIDIEQQVEYLSAKPRYLLVTDYDGVLAK
                     DTKTLEALDVKFEELPQYFDFFLAWKGIEKVEFEKENPADIKAAERFARIYDVLRKEN
                     NIIETNRGLDLFLIRLLFCFFAEDTDIFKRNSFTNLIKTLTEEDGSNLNKLFADLFIV
                     LDKNERDDVPSYLKEFPYVNGQLFTEPHTELEFSAKSRKLIIECGELLNWAKINPDIF
                     GSMIQAVASEESRSYLGMHYTSVPNIMKVIKPLFLDKLNQSFLDAYDDYTKLENLLTR
                     IGKIKFFDPACGSGNFLIITYKELRRMEINIIKRLQELLGEYLYVPSVTLSQFYGIEI
                     EDFAHDVAKLSLWIAEHQMNEELKNEVHNAVRPTLPLHTAGDIRCANAIRVEWTEVCP
                     AQGSEEVYVFGNPPYLGSKKQNKEHKSDMLSIFGKVKNGKMLDYISAWFYFGAKYAST
                     TNAKVAFVSTNSVTQGEQVSILWNELFKFGIQINFANKSFKWANNAKNNAAVIVVIVG
                     FGPLDTKVNKYLFVDETKKLVSNISPYLTDGENILVSSRTKPISDLPKLHFGNMPNDG
                     GGLLFTITEYTDAINKYPELVPYFKKFIGSVEFINGGLRYCLWLNEAKYEKIKSNPLI
                     QERISISKNHREKSTDKGTNKLALTPWKFRDTHETTNYSIVVPSVSSENRFYIPMGLA
                     GADTILSNLIYVIYDAEIYLLGILMSRMHMTWVKAVAGRLKTDYRYSAGLCYNTFPIP
                     ELSTRRKNEIEEAILEILDLREEQGGTLAELYNPSTMPIELKVAHEKLDGIVERAYRQ
                     KQFESDEERLEVLLKLYQEMTER"
RBS                  4152. .4161
                     /gene="yeeB"
gene                 4152. .5967
                     /gene="yeeB"
CDS                  4171. .5967
                     /gene="yeeB"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YeeB"
                     /protein_id="AAB66475.1"
                     /db_xref="GI:2318062"
                     /translation="MENIIEINYNQTGKSKKTNEYGMREMQARAFEKRNSQYLLVKAP
                     PASGKSRALMFIGLDKLINQGLRKVIVAVPERSIGSSFKNTDLKSYGFFENWKVDPRN
                     NLTVGGDNSKVKSFVRFMESDDQVLICTHSTLRFAFEKIDDKAFDNCLLAIDEFHHVS
                     ADVNSRLGELLRSIIHNSSAHIVAMTGSYFRGDSVPILLPEDEELFDKVTYSYYEQLD
                     GYEYLKGFGIGYHIYQGQYTSAINEVLDTNKKTIVHIPNVNSGESTKDKYDEVGKILD
                     MIGEVEYQDDDTGIIYVKRHSDGKTLKIADLVDDQVGRENVVAYLRNIEALDDLDIII
                     ALGMAKEGFDWPFCEHTLTVGYRGSLTEIVQIIGRCTRDSYNKTYAQFTNLIAMPDAK
                     DEVVTYTVNTMLKAISASLLMEQVLTPDFKFKRRRNESEKSSTTGELFVKGLKEPSTE
                     NVKKIIENDINDLKAKTCKIHKSKTFSGEVDPKVLNKVLIPKVIQKVYPNLSDEEIEE
                     VRQHVVLDTVMKGAKSEVVGSKEFIRMADKFINIEDLDINLIDSINPFQKAFEVLSKE
                     LNSPVLKLIQETIDAKRINFDEDELVFIWPKV"
RBS                  6064. .6069
                     /gene="yeeC"
gene                 6064. .7228
                     /gene="yeeC"
CDS                  6080. .7228
                     /gene="yeeC"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YeeC"
                     /protein_id="AAB66476.1"
                     /db_xref="GI:2318063"
                     /translation="MASNRYHSINEIMESQLFYQITTPPKKTQKAQYDPEVEKFLEII
                     EFVKENGREPQKVPTDLTERSLASRLIGIRKDPDRMEYLKEYDEIGLLEVRSKELTIP
                     KISSIDDILKSGLSALLGDNLNNDITRSIFDTSSLQKVTTMPEYVAKRKKIKDFGKFE
                     ELFKKCHKEITEGKRKILTFKNEQDIQSNSFYILKGVLLYVEDVGERKKAKGKTNARL
                     RCIFENGTESDMLLRSLSAELYKHGRRVTDNEDTLLDNVREDDVSTGFIYVLKSLSTD
                     PQISSIKNLYKIGFTTGSVENRIRNAENQSTYLYAPVEIVTTYQVFNMNASKFETAIH
                     HALENNNLDVSILGANGKMLVPKEWFVVTLEDLQAVIDEIVMMVHLYD"
gene                 complement(7490. .8180)
                     /gene="yeeD"
CDS                  complement(7490. .7795)
                     /gene="yeeD"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YeeD"
                     /protein_id="AAB66477.1"
```

```
                /db_xref="GI:2318064"
                /translation="MTVRKELIKQINLTITVIKTINQKNPTPMVKNILKRYEEAKEFI
                LQSSDGKFEEDLSKVRNKLDTLTRAYLESANDYMNPMLKEMHKTEKLLKEYDETTQS"
     RBS        complement(8175. .8180)
                /gene="yeeD"
     terminator 8277. .8298
                /note="rho independent terminator; deltaG= -10.2kcal/mol"
                /citation=[2]
     gene       complement(8308. .9470)
                /gene="yeeE"
     CDS        complement(8308. .9417)
                /gene="yeeE"
                /citation=[2]
                /codon_start=1
                /transl_table=11
                /product="YeeE"
                /protein_id="AAB66478.1"
                /db_xref="GI:2318065"
                /translation="MPAGVPYNAVNGVTLKEGLQKFAKVILPKPYGTSSSGRRTPAPH
                VPPVTVKYGEHFARWSRKKVLKPNIIYKTKEGYTYTTDNYGRITSVKADLQLGEAKRN
                QYAQTNAGKPQDRKPDDDGGHLIATQFKGSGQFDNIVPMNSQINRSGGKWYEMEQEWA
                KALSKKPPKKVAVQIEPVYSGDSLRPSYFDVTYKIGSRKEISVSIKISLGVRRMETRK
                MQDLYQLIGEKLNDIIPGEWTKIYLYAEVLDDSTMVLFHFRTPENNQIIYSQDIPSHY
                NVSKDIFKTLLRELRELFEELRTEHRNNNDDVWTNLTLTLDRSGEFQLDYNYDDILAS
                ELDGYERIAIWEYKNLGILPEDEDDKEFVISYLGL"
     RBS        complement(9464. .9470)
                /gene="yeeE"
     gene       complement(9541. .10791)
                /gene="yeeF"
     CDS        complement(9541. .10791)
                /gene="yeeF"
                /note="similar to Bacillus subtilis orf94 hypothetical
                59.7 kDA protein in CWLA-CISA intergenic region: SwissProt
                Accession Number P45942"

Query Match             48.1%;  Score 22.6;  DB 1;  Length 18622;
Best Local Similarity   68.9%;  Pred. No. 3.7e+02;
Matches   31;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       3 ctataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
           | | |||| || |  |||| | || ||| | ||||||| ||  ||
Db   17005 CCAAAATGAGGTCCAATATGCAGCATGAGCAGCCGCAAACTCTGA 16961


RESULT  12
AC009915
LOCUS       AC009915   101962 bp    DNA             HTG       10-FEB-2000
DEFINITION  Drosophila melanogaster chromosome 3 clone BACR09L14 (D751) RPCI-98
            09.L.14 map 94A-94A strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 84 unordered pieces.
ACCESSION   AC009915
VERSION     AC009915.4  GI:6957926
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 101962)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 101962)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Feb 11, 2000 this sequence version replaced gi:6598737.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 84 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1017: contig of 1017 bp in length
            *     1018     1097: gap of unknown length
            *     1098     1838: contig of 741 bp in length
            *     1839     1918: gap of unknown length
            *     1919     2723: contig of 805 bp in length
            *     2724     2803: gap of unknown length
            *     2804     3804: contig of 1001 bp in length
            *     3805     3884: gap of unknown length
            *     3885     4499: contig of 615 bp in length
            *     4500     4579: gap of unknown length
            *     4580     5031: contig of 452 bp in length
            *     5032     5111: gap of unknown length
            *     5112     6264: contig of 1153 bp in length
            *     6265     6344: gap of unknown length
            *     6345     6971: contig of 627 bp in length
            *     6972     7051: gap of unknown length
            *     7052     7722: contig of 671 bp in length
            *     7723     7802: gap of unknown length
            *     7803     8423: contig of 621 bp in length
            *     8424     8503: gap of unknown length
            *     8504     8988: contig of 485 bp in length
            *     8989     9068: gap of unknown length
            *     9069     9639: contig of 571 bp in length
            *     9640     9719: gap of unknown length
            *     9720    10396: contig of 677 bp in length
            *    10397    10476: gap of unknown length
            *    10477    11076: contig of 600 bp in length
            *    11077    11156: gap of unknown length
            *    11157    11722: contig of 566 bp in length
            *    11723    11802: gap of unknown length
            *    11803    13112: contig of 1310 bp in length
            *    13113    13192: gap of unknown length
            *    13193    14134: contig of 942 bp in length
            *    14135    14214: gap of unknown length
            *    14215    14808: contig of 594 bp in length
            *    14809    14888: gap of unknown length
            *    14889    15796: contig of 908 bp in length
            *    15797    15876: gap of unknown length
            *    15877    16944: contig of 1068 bp in length
            *    16945    17024: gap of unknown length
            *    17025    18221: contig of 1197 bp in length
            *    18222    18301: gap of unknown length
            *    18302    19185: contig of 884 bp in length
            *    19186    19265: gap of unknown length
            *    19266    19953: contig of 688 bp in length
            *    19954    20033: gap of unknown length
            *    20034    21120: contig of 1087 bp in length
            *    21121    21200: gap of unknown length
            *    21201    22120: contig of 920 bp in length
            *    22121    22200: gap of unknown length
```

```
*    22201    23445: contig of 1245 bp in length
*    23446    23525: gap of unknown length
*    23526    25145: contig of 1620 bp in length
*    25146    25225: gap of unknown length
*    25226    26308: contig of 1083 bp in length
*    26309    26388: gap of unknown length
*    26389    27288: contig of 900 bp in length
*    27289    27368: gap of unknown length
*    27369    28560: contig of 1192 bp in length
*    28561    28640: gap of unknown length
*    28641    29587: contig of 947 bp in length
*    29588    29667: gap of unknown length
*    29668    31197: contig of 1530 bp in length
*    31198    31277: gap of unknown length
*    31278    32671: contig of 1394 bp in length
*    32672    32751: gap of unknown length
*    32752    34749: contig of 1998 bp in length
*    34750    34829: gap of unknown length
*    34830    35733: contig of 904 bp in length
*    35734    35813: gap of unknown length
*    35814    36666: contig of 853 bp in length
*    36667    36746: gap of unknown length
*    36747    38088: contig of 1342 bp in length
*    38089    38168: gap of unknown length
*    38169    39885: contig of 1717 bp in length
*    39886    39965: gap of unknown length
*    39966    41304: contig of 1339 bp in length
*    41305    41384: gap of unknown length
*    41385    42989: contig of 1605 bp in length
*    42990    43069: gap of unknown length
*    43070    44998: contig of 1929 bp in length
*    44999    45078: gap of unknown length
*    45079    47171: contig of 2093 bp in length
*    47172    47251: gap of unknown length
*    47252    49068: contig of 1817 bp in length
*    49069    49148: gap of unknown length
*    49149    50817: contig of 1669 bp in length
*    50818    50897: gap of unknown length
*    50898    52445: contig of 1548 bp in length
*    52446    52525: gap of unknown length
*    52526    54327: contig of 1802 bp in length
*    54328    54407: gap of unknown length
*    54408    56212: contig of 1805 bp in length
*    56213    56292: gap of unknown length
*    56293    58084: contig of 1792 bp in length
*    58085    58164: gap of unknown length
*    58165    60628: contig of 2464 bp in length
*    60629    60708: gap of unknown length
*    60709    63251: contig of 2543 bp in length
*    63252    63331: gap of unknown length
*    63332    65404: contig of 2073 bp in length
*    65405    65484: gap of unknown length
*    65485    68291: contig of 2807 bp in length
*    68292    68371: gap of unknown length
*    68372    71873: contig of 3502 bp in length
*    71874    71953: gap of unknown length
*    71954    75078: contig of 3125 bp in length
*    75079    75158: gap of unknown length
*    75159    78908: contig of 3750 bp in length
*    78909    78988: gap of unknown length
*    78989    82484: contig of 3496 bp in length
*    82485    82564: gap of unknown length
*    82565    82854: contig of 290 bp in length
*    82855    82934: gap of unknown length
*    82935    83524: contig of 590 bp in length
*    83525    83604: gap of unknown length
*    83605    84401: contig of 797 bp in length
*    84402    84481: gap of unknown length
*    84482    85032: contig of 551 bp in length
*    85033    85112: gap of unknown length
*    85113    85692: contig of 580 bp in length
*    85693    85772: gap of unknown length
*    85773    86368: contig of 596 bp in length
*    86369    86448: gap of unknown length
*    86449    86881: contig of 433 bp in length
*    86882    86961: gap of unknown length
*    86962    87700: contig of 739 bp in length
*    87701    87780: gap of unknown length
*    87781    88521: contig of 741 bp in length
*    88522    88601: gap of unknown length
*    88602    89131: contig of 530 bp in length
*    89132    89211: gap of unknown length
*    89212    89797: contig of 586 bp in length
*    89798    89877: gap of unknown length
*    89878    90525: contig of 648 bp in length
*    90526    90605: gap of unknown length
*    90606    91446: contig of 841 bp in length
*    91447    91526: gap of unknown length
*    91527    91914: contig of 388 bp in length
*    91915    91994: gap of unknown length
*    91995    92444: contig of 450 bp in length
*    92445    92524: gap of unknown length
*    92525    93055: contig of 531 bp in length
*    93056    93135: gap of unknown length
*    93136    93647: contig of 512 bp in length
*    93648    93727: gap of unknown length
*    93728    94317: contig of 590 bp in length
*    94318    94397: gap of unknown length
*    94398    95262: contig of 865 bp in length
*    95263    95342: gap of unknown length

  Query Match              48.1%;  Score 22.6;  DB 55;  Length 101962;
  Best Local Similarity    68.9%;  Pred. No. 3e+02;
  Matches   31;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       1 agctataatgcggccgcatattcggcctgatcggccgcaagcttg 45
           ||| || | |||  ||| | ||||||| | || |  | ||| |||
Db   62414 AGCGATCAGGCGTTCGCCTGTTCGGCCAGCTCCGTGGAAAGTTTG 62458


RESULT  13
BSUB0004/c
LOCUS       BSUB0004   213190 bp    DNA             BCT       26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 4 of 21): from 600701 to
            813890.
ACCESSION   Z99107 AL009126
VERSION     Z99107.1  GI:2632866
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 213190)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
```

```
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
  TITLE     The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
REFERENCE   2  (bases 1 to 213190)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1. .213190
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            33. .740
                     /gene="ydfS"
     CDS             33. .740
                     /gene="ydfS"
                     /function="unknown"
                     /note="similar to hypothetical proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12373.1"
                     /db_xref="GI:2632867"
                     /db_xref="SPTREMBL:P96697"
                     /translation="MIELEVVIRTVASFGLLLIAERILGKQTISQMTIFDFIAAITLG
                     AIAAGLAYNTSIKPHNMAISFSIFVLTIFLISFLSIKNRKLRKFFAGDPTVLIQNGKI
                     LESNMRKMRYTLDYLNQQLREKEIFNIEEVLFAILETNGQLTVLRKPQFRHVTKQDLM
                     IAVNQEQRLPIELIMDGEIIENNLKQNRLTESWLLEELRKRDIKVKETVYAVLLGNGD
                     IYVDQYKDHISVPMDKE"
gene                 complement(755. .1186)
                     /gene="ydfT"
CDS                  complement(755. .1186)
                     /gene="ydfT"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12374.1"
                     /db_xref="GI:2632868"
                     /db_xref="SPTREMBL:P96698"
                     /translation="MDFEKIRKWLEITNEYKQSDFWTNVLKYKAPEHFFDSEASTFVY
                     DFYQDEEYNFIIVEMPGVYEEELTIRLLSKTQLLIKGTITPVFPAEMEVLRERYYGEI
                     ERIIQLPEAAETHLLQIQLLNGLLHISYPRQVETVAFNKGL"
gene                 complement(1199. .1441)
                     /gene="ydgA"
CDS                  complement(1199. .1441)
                     /gene="ydgA"
                     /function="unknown"
                     /note="similar to hypothetical proteins from B. subtilis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12375.1"
                     /db_xref="GI:2632869"
                     /db_xref="SPTREMBL:P96699"
                     /translation="MPYQINIANIKINGVTQNGNIDVGPTVHNSHTANSKYFGANFSL
                     GDLSPTSSLLNTGNIDSDVSDQDQIGNPSAPISNQI"
gene                 complement(1455. .1727)
                     /gene="ydgB"
CDS                  complement(1455. .1727)
                     /gene="ydgB"
                     /function="unknown"
                     /note="similar to hypothetical proteins from B. subtilis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12376.1"
                     /db_xref="GI:2632870"
                     /db_xref="SPTREMBL:P96700"
                     /translation="MPSTVINLYYLKINSISGNGSITIGEAAYNSPTNNQKSQGTNSS
                     FGDTSPTESVMENFLNDPDVNDQTSIGNSDTSNVNAPPIAPPPILD"
gene                 2026. .2613
                     /gene="ydgC"
CDS                  2026. .2613
                     /gene="ydgC"
                     /function="unknown"
                     /note="similar to hypothetical proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12377.1"
                     /db_xref="GI:2632871"
                     /db_xref="SPTREMBL:P96701"
                     /translation="MKSKGVESRKRLLKAAANEFSVRGFHDAKVSEIVKKAGFTQPSF
                     YLYFQSKEAIFAELITDFHSRVRKLTESLLLENGLNTEDVSKRVLLAVETVFQFLDED
                     KDLTKIGFFLNPEAKQMKKDLAMVLKENLEAEQRLGYFHSELDMETVAECLVGMIEHL
                     TESFLLTGIKDPASLAAEVVNLLIYGMLPKGNDVR"
gene                 join(2610. .2954,3074. .3091)
                     /gene="ydgD"
CDS                  2610. .2954
                     /gene="ydgD"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12378.1"
                     /db_xref="GI:2632872"
                     /db_xref="SPTREMBL:P96702"
                     /translation="MISIMMKVSLAVFMLAGGIIKVSRVPFQVEHWRHYQYPLWFLTV
                     TGILEIAGALAMTAGIWNRYAAIGAGVLFVVLMAGAIHAHMFRARQSVIMAIQAMICL
                     IVSIMIIMGSYT"
terminator           3074. .3091
                     /gene="ydgD"
gene                 3117. .3590
                     /gene="ydgE"
CDS                  3117. .3590
                     /gene="ydgE"
                     /function="unknown"
                     /note="similar to hypothetical proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12379.1"
                     /db_xref="GI:2632873"
                     /db_xref="SPTREMBL:P96703"
                     /translation="MAFTLEDMTEEEFEAFRGMSVQNFAKQNITSGTWTEKEAFEKSE
                     QAYENMIPNGRDSSNHYFWNITNDQGERMGWLWLYADPLHPQKEAFIYSFGLYEAFRG
                     KGLAQLALQTLDERARELGAERLALHVFAHNETAVYLYQKMGYAMTNIRMRKQLC"
terminator           complement(3705. .3722)
                     /gene="expZ"
gene                 complement(3705. .5393)
                     /gene="expZ"
CDS                  complement(3750. .5393)
                     /gene="expZ"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP-binding transport protein"
                     /protein_id="CAB12380.1"
                     /db_xref="GI:2632874"
                     /db_xref="SWISS-PROT:P39115"
                     /translation="MKEIVTLTNVSYEVKDQTVFKHVNASVQQGDIIGIIGKNGAGKS
                     TLLHLIHNDLAPAQGQILRKDIKLALVEQETAAYSFADQTPAEKKLLEKWHVPLRDFH
                     QLSGGEKLKARLAKGLSEDADLLLLDEPTNHLDEKSLQFLIQQLKHYNGTVILVSHDR
                     YFLDEAATKIWSLEDQTLIEFKGNYSGYMKFREKKRLTQQREYEKQQKMVERIEAQMN
                     GLASWSEKAHAQSTKKEGFKEYHRVKAKRTDAQIKSKQKRLEKELEKAKAEPVTPEYT
                     VRFSIDTTHKTGKRFLEVQNVTKAFGERTLFKNANFTIQHGEKVAIIGPNGSGKTTLL
                     NIILGQETAEGSVWVSPSANIGYLTQEVFDLPLEQTPEELFENETFKARGHVQNLMRH
                     LGFTAAQWTEPIKHMSMGERVKIKLMAYILEEKDVLILDEPTNHLDLPSREQLEETLS
                     QYSGTLLAVSHDRYFLEKTTNSKLVISNNGIEKQLNDVPSERNEREELRLKLETERQE
                     VLGKLSFMTPNDKGYKELDQAFNELTKRIKELDHQDKKD"
terminator           complement(5428. .5450)
```

```
                  /gene="ydgF"
     gene         complement(5428. .7089)
                  /gene="ydgF"
     terminator   complement(5678. .5703)
                  /gene="ydgF"
     CDS          complement(5713. .7089)
                  /gene="ydgF"
                  /function="unknown"
                  /note="similar to amino acid ABC transporter (permease)"
                  /codon_start=1
                  /transl_table=11
                  /protein_id="CAB12381.1"
                  /db_xref="GI:2632875"
                  /db_xref="SWISS-PROT:P96704"
                  /translation="MTDDMTKDNINQQTLQRGLKNRHIQLIAIGGAIGTGLFLGSGKS
                  IHFAGPSILFAYMITGIICFLIMRSLGELLLSNLNYHSFVDFVQDYLGDMAAFITGWT
                  YWFCWISIAMADLTAVGLYTQYWLPGVPQWVPGLIALIILLIMNLATVKLFGELEFWF
                  ALIKVIAILALIVIGLVMIFKGFSTSSGVSSFTNLWSHGGLFPNGMHGFILSFQMVVF
                  AFVGIELVGLTAGETENPEKVIPKAINNIPVRVLLFYIGALLVIMSIYPWDIINPSES
                  PFVQVFVAVGIVGAASIINFVVLTSAASACNSAVFSTSRMVYSLAKDHNAPESMAKLT

  Query Match             48.1%;  Score 22.6;  DB 2;  Length 213190;
  Best Local Similarity   68.9%;  Pred. No. 2.7e+02;
  Matches   31;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       3 ctataatgcggccgcatattcggcctgatcggccgcaagcttgga 47
           | | |||| || |  |||| | || ||| | |||||| ||  ||
Db 154579 CCAAAATGAGGTCCAATATGCAGCATGAGCAGCCGCAAACTCTGA 154535


RESULT  14
AE003738
LOCUS       AE003738   226159 bp    DNA             INV       05-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386035 section 63
            of 105, complete sequence.
ACCESSION   AE003738 AE002708
VERSION     AE003738.2  GI:10726692
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 226159)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 226159)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7300816.
FEATURES             Location/Qualifiers
     source          1. .226159
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
     mRNA            join(<18763. .21390,21641. .21907,21970. .>23157)
                     /gene="CG5732"
                     /product="CT18024"
                     /db_xref="FLYBASE:FBan0005732"
                     /db_xref="FLYBASE:FBgn0038934"
     gene            <18763. .>23157
                     /gene="CG5732"
                     /map="94A5-94A7"
                     /db_xref="FLYBASE:FBan0005732"
                     /db_xref="FLYBASE:FBgn0038934"
     CDS             join(18763. .21390,21641. .21907,21970. .23157)
                     /gene="CG5732"
                     /note="CG5732 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0005732"
                     /db_xref="FLYBASE:FBgn0038934"
                     /protein_id="AAF55959.2"
                     /db_xref="GI:10726693"
                     /translation="MAMAAAASSSAAAATATTTAISNSTNTTASPATTNTINTTTKTI
                     TTSEVPLEPASNMAPKALEIEDSGRNEPEDLEHGAAKPLEQRKTTVVHTCPRPPGGYK
                     YSMEFLYGIGSGMAGIPLNIPTPSSITPRTVRSTAPLLTTHMPLLTNMGVAPSRPSGI
                     RYPGSAGGSNGTTVAASTTPSTTVTTSSGSPGSGAEAITSSVSSSGLPAAGLQAAQFI
                     CQGYMPTGPQRRLWHAENAVWQFDRNYPYNQAYSPPYGIPMMPVGFEHPYGQRIIYPG
                     YYNQTPPGINPAAVAGLTRTNRPVTQQPHILTQPAVGTTESSEEAPATLGNAPQVSST
                     WRYGRPGTHRGRHAVPTAAAPALLHRDSKSFYNSIGSGMANGPRFKAPFVANVRNFQA
                     GAVATVAAGGAATTAVVGTSAPATGAASSSDQNVATKRNHQGAATQNNHRNRHNAKKG
                     GKNSVGKELTSNSSESLSNSSSKSQLNKRPSSSSSISPIKHPHRNYRNRMRYTATEPT
                     EQKAATTTVPISNYQPPQQSTSTVRRTSKFQGSNAYQAHTAAGRQQSRFYQSRHMDGY
                     VFQSGHYMVYAAGAPPVGLRPGKSPVSEATGAPPGAAAATAAAATAVAATAAAAEGGG
                     AAEGGGAAAAPVTASSATLEGEQPLDSDFDQRQEFADLGLDPANGGFSSDLEPNGIKQ
                     DTSELDTHSCLISHPNSEVDGDDNQSLASFAPSVESDDSDSELSDASVESVARDILVS
                     CLAVATGAEEPDFSGPNLVPYGDMHYLKELDKKNPPTNGYRSHRPYHQSHYAYHSQMS
                     PRGLSCCGDMLNQHSEDLVFKLDQNQPDGIESGKNIFLREITEQPDNISVASNLSCSP
                     SASSSKSVLAPMASKSNITMPEENNDDDELPLVVHNRYWREFFGYTPADRFLLRAKFV
                     EMRRPPKVMGCKNKWDPLSLSVWKKFLESQQTRHVYKIKMRLWRAIYTVAMKNYPRYG
                     LYLVGSSISYFGSKCSDMDICMLACTNPNIDSRMEAVYHLHVMKELLGRTNMFQDFNL
                     IEARVPILRFTDRCHKVEVDINFNNSVGIRNTHLLLYCYSQLDWRVRPMALTVKQWAQY
                     HNINNAKNMTISSYSLMLMVIHFLQVGASPPVLPCLHNLYPEKFGLLQPNDFGYVDMN
                     EVMAPYQSDNSQTLGDLLLSFLHYYSVFDYGKYAISIRVGGVLPIEVCRAATAPKNDI
                     HQWNELCIEEPFDQTNTARSVYDTDTFERIKTIFVASYRRLDSTRNLSAIFEDYDGPT
                     ILMQQPSVDSEIELYEGQHHRLLPNRGSSRSNSAIPSPRPSILMVDKATTAIWDDINN
                     KPDHPVLSHSNNYDATNECTGNGSLMGLKDNSVADKPPIA"
     mRNA            join(<84287. .84568,84948. .85222,85302. .85351,85825. .85931,
                     87395. .87481,88000. .88077,88145. .>88261)
                     /gene="CG7087"
                     /product="CT21905"
                     /db_xref="FLYBASE:FBan0007087"
                     /db_xref="FLYBASE:FBgn0038935"
     gene            <84287. .>88261
                     /gene="CG7087"
```

```
                /map="94A9-94A11"
                /db_xref="FLYBASE:FBan0007087"
                /db_xref="FLYBASE:FBgn0038935"
     CDS        join(84287. .84568,84948. .85222,85302. .85351,85825. .85931,
                87395. .87481,88000. .88077,88145. .88261)
                /gene="CG7087"
                /note="CG7087 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0007087"
                /db_xref="FLYBASE:FBgn0038935"
                /protein_id="AAF55960.1"
                /db_xref="GI:7300818"
                /translation="MAPFGDILATIERARQAAWAERSPLHSAAQVVGGSGGAGGSTGS
                GSGGAASGGSGGSSGVLHHRQQHGHRGHSMHGAGLPNSHSQPIYVPGKYSPSSCLSDK
                EEDEIYGFGYGVFAPRVARGGLTQQQQLLQQQTLQTQQSIQQQQQQMQQQQQQLPIVP
                GQQQQGPHQHQTLPPNVAHLNFVQQNCLSPRSAYFYEFPPTAEGRETKKRTTLARLLK
                GLKTVNRRDRNNQQNGAQARAANDRLRHFQMINGGAGGQQHSFEETIHRQQKQSHTRR
                VPQGKKDDDGEEEEEEERARSQTKNKQQNAINGDSSNRQQQQQQQQHLKQQQLQQ"
     mRNA       <92079. .>92330
                /gene="CG13864"
                /product="CT33378"
                /db_xref="FLYBASE:FBan0013864"
                /db_xref="FLYBASE:FBgn0040583"
                /evidence=not_experimental
     gene       <92079. .>92330
                /gene="CG13864"
                /map="94A10-94A11"
                /db_xref="FLYBASE:FBan0013864"
                /db_xref="FLYBASE:FBgn0040583"
                /evidence=not_experimental
     CDS        92079. .92330
                /gene="CG13864"
                /note="CG13864 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0013864"
                /db_xref="FLYBASE:FBgn0040583"
                /evidence=not_experimental
                /protein_id="AAG22167.1"
                /db_xref="GI:10726694"
                /translation="MYDEALRTGGGMGIAGLGMPLGVGGNGGGGGAAHYAVSALHHQG
                HWYDEPPYESDPDDFLMAGLNCGPAATIQVSLVLKILPK"
     mRNA       join(<96081. .96157,96208. .96277,96674. .>96901)
                /gene="CG13863"
                /product="CT33377"
                /db_xref="FLYBASE:FBan0013863"
                /db_xref="FLYBASE:FBgn0038936"
                /evidence=not_experimental
     gene       <96081. .>96901
                /gene="CG13863"
                /map="94A10-94A11"
                /db_xref="FLYBASE:FBan0013863"
                /db_xref="FLYBASE:FBgn0038936"
                /evidence=not_experimental
     CDS        join(96081. .96157,96208. .96277,96674. .96901)
                /gene="CG13863"
                /note="CG13863 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0013863"
                /db_xref="FLYBASE:FBgn0038936"
                /evidence=not_experimental
                /protein_id="AAF55961.1"
                /db_xref="GI:7300819"
                /translation="MIAKAATTRAPSQIYKALPPDSVRCRYGRLSNSFDISDSSKSRY
                AASECIGTNNCTARYRTLSGGIGESPSLSPSPSSDYEDIGVTRGHGCLPPSLLAAKAK
                KNGLPHGKANTICQKATVHHVS"
     mRNA       join(<98734. .98776,98909. .99054,99209. .99474,99546. .99667,
                100055. .100449,100525. .>100881)
                /gene="CG7086"
                /product="CT21903"
                /db_xref="FLYBASE:FBan0007086"
                /db_xref="FLYBASE:FBgn0038937"
     gene       <98734. .>100881
                /gene="CG7086"
                /map="94A10-94A12"
                /db_xref="FLYBASE:FBan0007086"
                /db_xref="FLYBASE:FBgn0038937"
     CDS        join(98734. .98776,98909. .99054,99209. .99474,99546. .99667,
                100055. .100449,100525. .100881)
                /gene="CG7086"
                /note="CG7086 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0007086"
                /db_xref="FLYBASE:FBgn0038937"
                /protein_id="AAF55962.1"
                /db_xref="GI:7300820"
                /translation="MSVSSNMHGWMFVHSTQALVRPGSPHSSISAEDRTSMASCICKA
```

```
  Query Match             48.1%;  Score 22.6;  DB 30;  Length 226159;
  Best Local Similarity   68.9%;  Pred. No. 2.7e+02;
  Matches   31;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       1 agctataatgcggccgcatattcggcctgatcggccgcaagcttg 45
           ||| || | |||  ||| | ||||||| | || |  | ||| |||
Db   97887 AGCGATCAGGCGTTCGCCTGTTCGGCCAGCTCCGTGGAAAGTTTG 97931


RESULT  15
AVIGLNA/c
LOCUS       AVIGLNA      1950 bp    DNA              BCT       06-MAR-1995
DEFINITION  A.vinelandii glutamine synthetase (glnA) gene, complete cds.
ACCESSION   M57275
VERSION     M57275.1  GI:142308
KEYWORDS    glutamine sythetase.
SOURCE      Azobacter vinelandii DNA.
  ORGANISM  Azotobacter vinelandii
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Azotobacter.
REFERENCE   1  (bases 1 to 1950)
  AUTHORS   Toukdarian,A., Saunders,G., Selman-Sosa,G., Santero,E., Woodley,P.
            and Kennedy,C.
  TITLE     Molecular analysis of the Azotobacter vinelandii glnA gene encoding
            glutamine synthetase
  JOURNAL   J. Bacteriol. 172 (11), 6529-6539 (1990)
  MEDLINE   91035268
FEATURES             Location/Qualifiers
     source          1. .1950
                     /organism="Azotobacter vinelandii"
                     /isolate="UW136"
                     /strain="OP"
                     /db_xref="taxon:354"
                     /clone="pLV50"
     mRNA            446. .>1950
                     /gene="glnA"
                     /evidence=experimental
     gene            446. .1950
                     /gene="glnA"
     CDS             478. .1881
                     /gene="glnA"
                     /EC_number="6.3.1.2"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamine synthetase"
                     /protein_id="AAA62673.1"
                     /db_xref="GI:142309"
                     /translation="MSKSLQLIKEHDVKWIDLRFTDTKGKQQHVTMPARDVDDDFFEY
                     GKMFDGSSIAGWKGIEASDMILMPDDSTAVLDPFTEEPTLIIVCDIIEPSTMQGYDRD
                     PRAIARRAEEYLKSTGIGDTAFFGPEPEFFIFDEVKYKSDISGSMFKIFSEQAAWNTD
                     ADFEGGNKGHRPGVKGGYFPVPPVDHDHEIRTAMCNALEEMGLKVEVHHHEVATAGQN
                     EIGVSFNTLVAKADEVQTLKYCVHNVADAYGKTVTFMPKPLYGDNGSGMHVHMSIAKD
                     GKNTFAGEGYAGLSDTALYFIGGIIKHGKALNGFTNPSTNSYKRLVPGFEAPVMLAYS
                     ARNRSASIRIPYVNSPKARRIEARFPDPSANPYLAFAALLMAGLDGIQNKIHPGDAAD
                     KNLYDLPPEEAKEIPQVCGSLKEALEELDKGRAFLTKGGVFSDDFIDAYLELKSEEEI
                     KVRTFVHPLEYDLYYSV"
BASE COUNT      424 a    646 c    543 g    337 t
ORIGIN
```

Query Match 47.2%; Score 22.2; DB 2; Length 1950;
Best Local Similarity 77.1%; Pred. No. 7e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy      2 gctataatgcggccgcatattcggcctgatcggcc 36
          |||||||| | ||| | | ||  |||||| |||||
Db    200 GCTATAATCCCGCCCCCTCTTTTGCCTGACCGGCC 166
```

Search completed: March 4, 2001, 12:04:02
Job time: 5617 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 09:48:30 ; Search time 1300.69 Seconds
(without alignments)
242.438 Million cell updates/sec

Title: US-09-101-423B-7
Perfect score: 45
Sequence: 1 aatccaagcttgcggccgat.....cgaatatgcggccgcattat 45

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*
44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estpl6:*
118: em_estpl7:*
119: em_estpl8:*
120: em_estro14:*
121: em_estro15:*
122: em_estro16:*
123: em_estro17:*
124: em_estro18:*
125: em_estro19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est75:*
136: gb_est76:*
137: gb_est77:*
138: gb_est78:*
139: gb_est79:*
140: gb_est80:*
141: gb_est81:*
142: gb_est82:*
143: gb_est83:*
144: gb_est84:*
145: gb_est85:*
146: gb_est86:*
147: gb_est87:*
148: gb_gss1:*
149: gb_gss2:*
150: gb_gss3:*
151: gb_gss4:*
152: em_gss1:*
153: em_gss2:*
154: em_gss3:*
155: em_gss4:*
156: gb_gss5:*
157: gb_gss6:*
158: gb_gss7:*
159: gb_gss8:*
160: gb_gss9:*
161: gb_gss10:*
162: gb_gss11:*
163: gb_gss12:*
164: gb_gss13:*
165: gb_gss14:*
166: gb_gss15:*
167: gb_gss16:*
168: gb_gss17:*
169: gb_gss18:*
170: gb_gss19:*
171: gb_gss20:*
172: gb_gss21:*
173: gb_gss22:*
174: gb_gss23:*
175: gb_gss24:*
176: em_gss5:*
177: em_gss6:*
178: em_gss7:*
179: em_gss8:*
180: em_gss9:*
181: em_gss10:*
182: em_gss11:*
183: em_gss12:*
184: em_gss13:*
185: em_gss14:*
186: em_gss15:*
187: em_gss16:*
188: em_gss17:*
189: em_gss18:*
190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	23.4	52.0	313	18	AI283439	AI283439 qh93e10.x
c	2	23.4	52.0	520	4	AA236621	AA236621 zs43g09.r
	3	23.2	51.6	625	21	AI514427	AI514427 LD40932.5
	4	23.2	51.6	1054	134	BE036773	BE036773 MP05D02 M
	5	22.4	49.8	705	97	AW977483	AW977483 EST389592
	6	22.2	49.3	528	134	BE062593	BE062593 QV1-BT026
	7	22.2	49.3	558	163	AQ841408	AQ841408 T136718b
	8	22	48.9	266	146	W06254	W06254 TgESTzy78d0
c	9	22	48.9	267	14	AA952587	AA952587 TENS1771
c	10	22	48.9	313	23	AI667900	AI667900 TENGO871
	11	22	48.9	373	146	W35559	W35559 TgESTzy89b0
c	12	21.8	48.4	395	87	AW185272	AW185272 se89c06.y
c	13	21.8	48.4	436	37	AV624476	AV624476 AV624476
	14	21.8	48.4	610	87	AW181098	AW181098 MgA0062r
	15	21.8	48.4	615	157	AQ399560	AQ399560 mgxb0014O
	16	21.8	48.4	635	157	AQ398650	AQ398650 mgxb0001B
	17	21.8	48.4	640	87	AW181091	AW181091 MgA0032r
c	18	21.6	48.0	541	159	AQ578848	AQ578848 nbxb0093F
c	19	21.6	48.0	650	7	AA438826	AA438826 LD13184.5
c	20	21.6	48.0	677	138	BE976052	BE976052 bs46f11.x
c	21	21.6	48.0	707	97	AW940801	AW940801 GH20963.3
	22	21.4	47.6	947	193	CNS05GOO	AL336561 Tetraodon
c	23	21.2	47.1	263	36	AV425024	AV425024 AV425024
c	24	21.2	47.1	266	36	AV426457	AV426457 AV426457
c	25	21.2	47.1	403	36	AV409533	AV409533 AV409533
c	26	21.2	47.1	463	36	AV422473	AV422473 AV422473
c	27	21.2	47.1	474	191	CNS03AWS	AL235765 Tetraodon
c	28	21.2	47.1	912	97	AW925378	AW925378 HVSMEg000
c	29	21.2	47.1	991	191	CNS02VLE	AL215915 Tetraodon
c	30	21	46.7	188	23	AI690000	AI690000 tx32a09.x
c	31	21	46.7	220	91	AW491261	AW491261 UI-M-BH3-
c	32	21	46.7	244	137	BE943769	BE943769 UI-M-BH3-
c	33	21	46.7	272	91	AW491371	AW491371 UI-M-BH3-
c	34	21	46.7	331	144	R87445	R87445 ym89b12.s1
c	35	21	46.7	350	15	AI024322	AI024322 ov67g07.x
	36	21	46.7	351	107	BE363902	BE363902 PI1_10_G0
	37	21	46.7	352	24	AI768069	AI768069 wi46d08.x
	38	21	46.7	372	14	AB009130	AB009130 AB009130
	39	21	46.7	391	36	AV420471	AV420471 AV420471
c	40	21	46.7	408	7	AA463976	AA463976 zx86f01.s
c	41	21	46.7	430	105	BE222661	BE222661 hu49h10.x
	42	21	46.7	448	110	BE647694	BE647694 UI-M-BH1-
	43	21	46.7	463	39	AW077673	AW077673 fj36b05.y
	44	21	46.7	487	18	AI292116	AI292116 qm88e02.x
	45	21	46.7	488	18	AI291238	AI291238 qm11b04.x

ALIGNMENTS

RESULT 1
AI283439/c
LOCUS AI283439 313 bp mRNA EST 23-NOV-1998
DEFINITION qh93e10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1854570 3' similar to TR:P76904 P76904 SIMILAR TO ;, mRNA sequence.
ACCESSION AI283439
VERSION AI283439.1 GI:3921672
KEYWORDS EST.

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 297.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1854570"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT       67 a     96 c    104 g     46 t
ORIGIN


  Query Match             52.0%;  Score 23.4;  DB 18;  Length 313;
  Best Local Similarity   73.2%;  Pred. No. 27;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      4 ccaagcttgcggccgatcaggccgaatatgcggccgcatta 44
          || ||||  | | ||| | ||||||||  | |||||||| ||
Db     75 CCCAGCTGCCCGGCGAACTGGCCGAATTGGTGGCCGCAGTA 35


RESULT  2
AA236621/c
LOCUS       AA236621      520 bp    mRNA            EST       07-AUG-1997
DEFINITION  zs43g09.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:688000 5'
            similar to TR:G1079677 G1079677 LPE5P. ;, mRNA sequence.
ACCESSION   AA236621
VERSION     AA236621.1  GI:1860641
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 618   Std Error: 0.00
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 425.
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Homo sapiens"
                     /db_xref="GDB:5592156"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:688000"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT      109 a    170 c    149 g     92 t
ORIGIN


  Query Match             52.0%;  Score 23.4;  DB 4;  Length 520;
  Best Local Similarity   73.2%;  Pred. No. 29;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      3 tccaagcttgcggccgatcaggccgaatatgcggccgcatt 43
          |||| ||  ||||| |  ||||| |||  |||||| | |||
Db     62 TCCATGCCAGCGGCAGTCCAGGCGGAACTTGCGGCAGAATT 22


RESULT  3
AI514427
LOCUS       AI514427      625 bp    mRNA            EST       16-MAR-1999
DEFINITION  LD40932.5prime LD Drosophila melanogaster embryo pOT2 Drosophila
            melanogaster cDNA clone LD40932 5prime similar to U13637: yl
            FBgn0004649 PID:g535346 SWISS-PROT:P98163, mRNA sequence.
ACCESSION   AI514427
VERSION     AI514427.1  GI:4418489
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 625)
  AUTHORS   Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein
            ,P., Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 409  row: C  column: 8
            High quality sequence stop: 507.
FEATURES             Location/Qualifiers
     source          1. .625
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
```

```
                     /clone="LD40932"
                     /clone_lib="LD Drosophila melanogaster embryo pOT2"
                     /sex="male and female"
                     /dev_stage="0 to 24 hours mixed stage embryonic"
                     /lab_host="XL1 Blue"
                     /note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2. "
BASE COUNT      138 a    166 c    184 g    137 t
ORIGIN


  Query Match             51.6%;  Score 23.2;  DB 21;  Length 625;
  Best Local Similarity   70.5%;  Pred. No. 35;
  Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      2 atccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
          |||||||| ||| |       || || || ||||||||||| |
Db    306 ATCCAAGCGTGCAGTTTCGTGGGACGCATGTGCGGCCGCATTGT 349


RESULT   4
BE036773
LOCUS       BE036773      1054 bp    mRNA            EST       07-JUN-2000
DEFINITION  MP05D02 MP Mesembryanthemum crystallinum cDNA 5', mRNA sequence.
ACCESSION   BE036773
VERSION     BE036773.1  GI:8331789
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
            Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 1054)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu
            Best blastx match: 'emb|CAB38794.1| (AL035678) putative protein
            [Arabidopsis thaliana] 267 9e-71'.
            Insert Length: 1   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .1054
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone_lib="MP"
                     /tissue_type="apical meristem and leaf primordia"
                     /dev_stage="6 weeks"
                     /note="3 d 500mM NaCl"
BASE COUNT      259 a    247 c    244 g    299 t      5 others
ORIGIN


  Query Match             51.6%;  Score 23.2;  DB 134;  Length 1054;
  Best Local Similarity   70.5%;  Pred. No. 37;
  Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      2 atccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
          ||| ||| ||||||| ||| ||   | ||  | ||| |||| ||
Db    553 ATCGAAGGTTGCGGCTGATAAGATTGCATTAGAGGCTGCATCAT 596


RESULT   5
AW977483
LOCUS       AW977483      705 bp    mRNA            EST       02-JUN-2000
DEFINITION  EST389592 MAGE resequences, MAGO Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW977483
VERSION     AW977483.1  GI:8168734
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 378
            Seq primer: Forward.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGO"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      175 a    187 c    147 g    196 t
ORIGIN


  Query Match             49.8%;  Score 22.4;  DB 97;  Length 705;
  Best Local Similarity   72.5%;  Pred. No. 74;
  Matches   29;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      6 aagcttgcggccgatcaggccgaatatgcggccgcattat 45
          |||||||||||| ||||||| |  || | ||    |||
Db      1 AAGCTTGCGGCCGCTCAGGCCCCACCTGAGACCTATCTAT 40


RESULT   6
BE062593
LOCUS       BE062593      528 bp    mRNA            EST       09-JUN-2000
DEFINITION  QV1-BT0260-281099-023-c01 BT0260 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE062593
VERSION     BE062593.1  GI:8407243
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
```

(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV1-BT0260-281099-023-c01&t3=1999-10-28&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 9
High quality sequence stop: 426.

FEATURES Location/Qualifiers
source 1. .528
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0260"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 101 a 180 c 152 g 94 t 1 others
ORIGIN

Query Match 49.3%; Score 22.2; DB 134; Length 528;
Best Local Similarity 77.1%; Pred. No. 86;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy      9 cttgcggccgatcaggccgaatatgcggccgcatt 43
          ||||| |||| ||  || |||  |||||||| |||
Db    298 CTTGCTGCCGTTCTCGCGGAAGGTGCGGCCGAATT 332
```

RESULT 7
AQ841408
LOCUS AQ841408 558 bp DNA GSS 27-SEP-1999
DEFINITION T136718b shotgun sub-library of BAC clone 16K14 Medicago truncatula genomic clone 16-K-14-022, DNA sequence.
ACCESSION AQ841408
VERSION AQ841408.1 GI:5930433
KEYWORDS GSS.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE 1 (bases 1 to 558)
AUTHORS Kim,D., Choi,H., Peng,H., Ellis,L. and Cook,DR.
TITLE BAC survey sequencing of Medicago truncatula (1999c)
JOURNAL Unpublished (1999)
COMMENT Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F. Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@ppserver.tamu.edu
Other name: BSC-19-22; date: 6/8/99; Submitted to the Database of Genome Survey Sequences (GSS) on 09/27/99; More information is available at 'http://chrysie.tamu.edu/medicago'.
Seq primer: pUC-C
Class: BAC subclone.

FEATURES Location/Qualifiers
source 1. .558
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="16-K-14-022"
/clone_lib="shotgun sub-library of BAC clone 16K14"
/note="Vector: pUC18; BAC survey sequences were obtained from sheared BAC DNA subcloned into the SmaI site of pUC18. The template DNA for sequencing was obtained by PCR using universal primers. Sequencing reactions were primed from the PUC-C primer site (CAGGAAACAGCTATGACCATGATTACGA) in the pUC18 polylinker."

BASE COUNT 168 a 76 c 95 g 219 t
ORIGIN

Query Match 49.3%; Score 22.2; DB 163; Length 558;
Best Local Similarity 69.8%; Pred. No. 86;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
Qy      3 tccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
          | | |||||| ||| |  |||||   ||| | ||  |||||||
Db    444 TGCTAGCTTGTGGCTGCACAGGCTTTATAAGGGGAAGCATTAT 486
```

RESULT 8
W06254
LOCUS . W06254 266 bp mRNA EST 22-MAY-2000
DEFINITION TgESTzy78d04.r1 TgRH Tachyzoite cDNA Toxoplasma gondii cDNA clone tgzy78d04.r1 5', mRNA sequence.
ACCESSION W06254
VERSION W06254.1 GI:1278977
KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae; Toxoplasma.
REFERENCE 1 (bases 1 to 266)
AUTHORS Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioka,J.A., Aslett,M.A., Dietrich,N., Dubuque,T., Hillier,L., Kucaba,T., Wan,K.L., Waterston,R.H. and Boothroyd,J.
TITLE WashU-Merck-Stanford-NIH Toxoplasma EST project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
David Sibley at toxoest@borcim.wustl.edu for further information relating to organism, clone or library availability.
Seq primer: T3
High quality sequence stop: 229.

FEATURES Location/Qualifiers
source 1. .266
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"
/clone="tgzy78d04.r1"
/clone_lib="TgRH Tachyzoite cDNA"
/lab_host="XL1-Blue MRF'"
/note="Vector: Lambda ZAP; Site_1: EcoRI; Site_2: XhoI; Toxoplasma RH strain tachyzoites were grown in human foreskin fibroblast cultures in vitro. The library was constructed by K.L. Wan, Cambridge University. cDNAs were synthesized from polyA RNAs by oligo d(T) priming and directionally cloned into the EcoRI to XhoI sites of the Lambda ZapII vector using the ZAP-cDNA synthesis kit (Stratagene). WARNING: the library contains a small percentage of cDNAs derived from the human host cells."

BASE COUNT 36 a 66 c 78 g 83 t 3 others
ORIGIN

Query Match 48.9%; Score 22; DB 146; Length 266;
Best Local Similarity 73.7%; Pred. No. 96;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy      1 aatccaagcttgcggccgatcaggccgaatatgcggcc 38
          ||||||  ||||| |  |||| ||  ||  ||||||||
Db     81 AATCCACACTTGCCGAAGATCCGGTGGACGATGCGGCC 118
```

```
RESULT   9
AA952587/c
LOCUS       AA952587      267 bp    mRNA            EST       29-OCT-1998
DEFINITION  TENS1771 T. cruzi epimastigote normalized cDNA Library Trypanosoma
            cruzi cDNA clone 1771 5', mRNA sequence.
ACCESSION   AA952587
VERSION     AA952587.1  GI:3115683
KEYWORDS    EST.
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 267)
  AUTHORS   Verdun,R.E., Di Paolo,N.C., Urmenyi,T.P., Rondinelli,E., Frasch
            ,A.C.C. and Sanchez,D.O.
  TITLE     Gene discovery through expressed sequence tag sequencing in
            trypanosoma cruzi
  JOURNAL   Infect. Immun. 66 (11), 5393-5398 (1998)
  MEDLINE   99003155
COMMENT     Contact: Sanchez D.O.
            Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
            San Martin)
            Av. Gral Paz entre Albarellos y Constituyentes, INTI edificio 24
            cp(1650) San Martin, Prov. de BS AS. Argentina
            Tel: (54-1)752-9639 or (54-1)752-0021
            Fax: (54-1)752-0021 or (54-1)752-9639
            Email: dsanchez@inti.gov.ar
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .267
                     /organism="Trypanosoma cruzi"
                     /strain="Cl-Brenner"
                     /db_xref="taxon:5693"
                     /clone="1771"
                     /clone_lib="T. cruzi epimastigote normalized cDNA Library"
                     /cell_type="epimastigote"
                     /note="cDNA library constructed with oligo dT primed
                     epimastigote mRNA and cloned in pt7t318D phagemid with
                     modified polylinker (PHARMACIA)"
BASE COUNT       66 a     79 c     72 g     46 t      4 others
ORIGIN


  Query Match             48.9%;  Score 22;  DB 14;  Length 267;
  Best Local Similarity   73.7%;  Pred. No. 96;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       6 aagcttgcggccgatcaggccgaatatgcggccgcatt 43
           |||||||||||||  | ||| | |  |||||| | | |
Db     237 AAGCTTGCGGCCGCACGGGCTGCAGTTGCGGCAGTACT 200


RESULT  10
AI667900/c
LOCUS       AI667900      313 bp    mRNA            EST       07-JUL-1999
DEFINITION  TENGO871 T. Cruzi epimastigote normalised cDNA Library Trypanosoma
            cruzi cDNA clone n442.r 5', mRNA sequence.
ACCESSION   AI667900
VERSION     AI667900.1  GI:4826272
KEYWORDS    EST.
SOURCE      Trypanosoma cruzi.
  ORGANISM  Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1  (bases 1 to 313)
  AUTHORS   Gonzalez Rey,E., Remisz,E., Delgado Garcia,A. and Gonzalez,A.
  TITLE     Characterization of ESTs from Trypanosoma cruzi epimastigotes
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Delgado Alberto
            Departamento de Biologia Molecular, Lab 303
            Instituto de Parasitologia y Biomedicina
            Consejo Superior de Investigaciones Cientificas C/ Ventanilla No 11
            , E-18001, Granada, Spain
            Tel: 34 958 805058
            Fax: 34 958 203323
            Similar to T.cruzi elongation factor 1-gamma (mammalian homologue)
            mRNA, gb|L17307|TRBEF1GHOM with Blastn.
            Seq primer: T7
            High quality sequence stop: 313.
FEATURES             Location/Qualifiers
     source          1. .313
                     /organism="Trypanosoma cruzi"
                     /strain="Cl - Brenner"
                     /db_xref="taxon:5693"
                     /clone="n442.r"
                     /clone_lib="T. Cruzi epimastigote normalised cDNA Library"
                     /cell_type="epimastigote"
                     /note="Site_1: EcoRI; Site_2: NotI; cDNA library
                     constructed with oligo dt primed  epimastigote mRNA and
                     cloned in pt7t318D  phagemid with modified polylinker"
BASE COUNT       72 a     91 c     83 g     67 t
ORIGIN


  Query Match             48.9%;  Score 22;  DB 23;  Length 313;
  Best Local Similarity   73.7%;  Pred. No. 97;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       6 aagcttgcggccgatcaggccgaatatgcggccgcatt 43
           |||||||||||||  | ||| | |  |||||| | | |
Db     255 AAGCTTGCGGCCGCACGGGCTGCAGTTGCGGCAGTACT 218


RESULT  11
W35559
LOCUS       W35559        373 bp    mRNA            EST       22-MAY-2000
DEFINITION  TgESTzy89b07.r1 TgRH Tachyzoite cDNA Toxoplasma gondii cDNA clone
            tgzy89b07.r1 5', mRNA sequence.
ACCESSION   W35559
VERSION     W35559.1  GI:1317476
KEYWORDS    EST.
SOURCE      Toxoplasma gondii.
  ORGANISM  Toxoplasma gondii
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 1 to 373)
  AUTHORS   Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioka,J.A., Aslett,M.A.
            , Dietrich,N., Dubuque,T., Hillier,L., Kucaba,T., Wan,K.L.,
            Waterston,R.H. and Boothroyd,J.
  TITLE     WashU-Merck-Stanford-NIH Toxoplasma EST project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M
            WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: toxo@watson.wustl.edu
            David Sibley at toxoest@borcim.wustl.edu for further information
            relating to organism, clone or library availability.
            Seq primer: T3
            High quality sequence stop: 313.
FEATURES             Location/Qualifiers
     source          1. .373
                     /organism="Toxoplasma gondii"
                     /strain="RH"
                     /db_xref="taxon:5811"
                     /clone="tgzy89b07.r1"
                     /clone_lib="TgRH Tachyzoite cDNA"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: Lambda ZAP; Site_1: EcoRI; Site_2: XhoI;
                     Toxoplasma RH strain tachyzoites were grown in human
                     foreskin fibroblast cultures in vitro. The library was
                     constructed by K.L. Wan, Cambridge University.  cDNAs were
                     synthesized from polyA RNAs by oligo d(T) priming and
                     directionally cloned into the EcoRI to XhoI sites of the
```

Lambda ZapII vector using the ZAP-cDNA synthesis kit (Stratagene). WARNING: the library contains a small percentage of cDNAs derived from the human host cells."
BASE COUNT 57 a 88 c 110 g 109 t 9 others
ORIGIN

Query Match 48.9%; Score 22; DB 146; Length 373;
Best Local Similarity 73.7%; Pred. No. 99;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy      1 aatccaagcttgcggccgatcaggccgaatatgcggcc 38
          ||||||  |||||  |  ||||  ||  ||  ||||||||
Db     74 AATCCACACTTGCCGAAGATCCGGTGGACGATGCGGCC 111
```

RESULT 12
AW185272/c
LOCUS AW185272 395 bp mRNA EST 19-NOV-1999
DEFINITION se89c06.y1 Gm-c1023 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1023-1859 5' similar to TR:P93194 P93194 RECEPTOR-LIKE PROTEIN KINASE. ;, mRNA sequence.
ACCESSION AW185272
VERSION AW185272.1 GI:6454589
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 395)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Putative full length read
vector to vector length is 396.
FEATURES Location/Qualifiers
source 1. .395
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1023-1859"
/clone_lib="Gm-c1023"
/tissue_type="seed coats of greenhouse grown plants"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This cDNA library was constructed from mRNA isolated from seed coats (100-200mgs) of greenhouse grown plants. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complimentary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restrictions site. Sal I linkers adapers were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. This library was constructed by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT 122 a 71 c 82 g 120 t
ORIGIN

Query Match 48.4%; Score 21.8; DB 87; Length 395;
Best Local Similarity 70.7%; Pred. No. 1.2e+02;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```
Qy      2 atccaagcttgcggccgatcaggccgaatatgcggccgcat 42
          |||||  |||||  |  |  |||  |||  |||  ||||    |  |||
Db    247 ATCCACGCTTGTGCCAGATAAGGAGGAAAATGCATTCTCAT 207
```

RESULT 13
AV624476/c
LOCUS AV624476 436 bp mRNA EST 11-OCT-2000
DEFINITION AV624476 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone LC078c03_r 5', mRNA sequence.
ACCESSION AV624476
VERSION AV624476.1 GI:10773653
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 436)
AUTHORS Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K., Nakamura,Y. and Tabata,S.
TITLE Generation of Expressed Sequence Tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. (2000) In press
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .436
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="LC078c03_r"
/clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"
BASE COUNT 69 a 157 c 116 g 94 t
ORIGIN

Query Match 48.4%; Score 21.8; DB 37; Length 436;
Best Local Similarity 78.8%; Pred. No. 1.2e+02;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy      7 agcttgcggccgatcaggccgaatatgcggccg 39
          ||||  |||  |  ||||  ||  ||||  ||||||||
Db    434 AGCTGGCGCGCCATCATGCGGAATGTGCGGCCG 402
```

RESULT 14
AW181098
LOCUS AW181098 610 bp mRNA EST 17-NOV-1999
DEFINITION MgA0062r MgA Library Mycosphaerella graminicola cDNA clone MgA0062 3' similar to OMEGA-6 FATTY ACID DESATURASE, ENDOPLASMIC RETICULUM ISOZYME 2, mRNA sequence.
ACCESSION AW181098
VERSION AW181098.1 GI:6448332
KEYWORDS EST.
SOURCE Mycosphaerella graminicola.

```
  ORGANISM  Mycosphaerella graminicola
            Eukaryota; Fungi; Ascomycota; Ascomycota incertae sedis;
            Mycosphaerellaceae; Mycosphaerella.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Keon,J.P.R., Bailey,A.M. and Hargreaves,J.A.
  TITLE     A group of expressed cDNA sequences from the wheat fungal leaf
            blotch pathogen, Mycosphaerella graminicola
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Hargreaves JA
            Cell Biology Department
            IACR-Long Ashton Research Station
            Long Ashton, Bristol, BS41 9AF, UK
            Tel: +44(0)1275 392181
            Fax: +44(0)1275 394281
            Email: john.hargreaves@bbsrc.ac.uk
            Insert Length: 2200   Std Error: 0.00
            Seq primer: M13 forward.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Mycosphaerella graminicola"
                     /strain="Strit"
                     /db_xref="taxon:54734"
                     /clone="MgA0062"
                     /clone_lib="MgA Library"
                     /note="Vector: pSPORT1; Library constructed from cultures
                     utilizing ammomium ions as a source of nitrogen"
BASE COUNT      125 a    169 c    156 g    159 t      1 others
ORIGIN


  Query Match             48.4%;  Score 21.8;  DB 87;  Length 610;
  Best Local Similarity   78.8%;  Pred. No. 1.2e+02;
  Matches   26;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       7 agcttgcggccgatcaggccgaatatgcggccg 39
           || |||||||||||   |||||||  |||| ||
Db     532 AGGTTGCGGCCGATGTAGCCGAATTCGCGGTCG 564


RESULT  15
AQ399560
LOCUS       AQ399560      615 bp    DNA             GSS       06-MAR-1999
DEFINITION  mgxb0014O23f CUGI Rice Blast BAC Library Pyricularia grisea genomic
            clone mgxb0014O23f, DNA sequence.
ACCESSION   AQ399560
VERSION     AQ399560.1  GI:4370587
KEYWORDS    GSS.
SOURCE      Pyricularia grisea.
  ORGANISM  Pyricularia grisea
            Eukaryota; Fungi; Ascomycota; Ascomycota incertae sedis;
            Magnaporthaceae; anamorphic Magnaporthaceae; Pyricularia.
REFERENCE   1  (bases 1 to 615)
  AUTHORS   Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,
            Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
            Genome
  JOURNAL   Unpublished (1998)
COMMENT.    Contact: Dean RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson Universiy, Clemson, SC 29634
            Tel: 864 656 5737
            Fax: 864 656 4293
            Email: rdean@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence start: 40
            High quality sequence stop: 464.
FEATURES             Location/Qualifiers
     source          1. .615
                     /organism="Pyricularia grisea"
                     /strain="70-15"
                     /db_xref="taxon:89476"
                     /clone="mgxb0014O23f"
                     /clone_lib="CUGI Rice Blast BAC Library"
                     /tissue_type="Protoplasts"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACWICH; Site_1: HindIII; Site_2: HindIII;
                     Rice blast is one of the most devestating fungal diseases
                     of rice  world wide.  It is a filamentous ascomycete with
                     a haploid genome (n=7)  of approximately 40 Mbp.  Rice
                     blast is an important model fungal pathogen  for studying
                     numerous aspects of the fungal-host interaction.   In
                     order to facilitate genome wide analysis, a BAC library
                     containing  9216 clones with an average insert size of 130
                     kbp was constructed.  This library represents greater
                     than 25X genome coverage.  High density  colony filters
                     are available upon request."
BASE COUNT      160 a    149 c    154 g    151 t      1 others
ORIGIN


  Query Match             48.4%;  Score 21.8;  DB 157;  Length 615;
  Best Local Similarity   70.7%;  Pred. No. 1.2e+02;
  Matches   29;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 aatccaagcttgcggccgatcaggccgaatatgcggccgca 41
             ||| | ||||||  |||| ||| | |||| ||| | | | |
Db      36 AATGCCAGCTTGGAGCCGCTCAAGGCGAAAATGGGCCAGGA 76


Search completed: March  4, 2001, 11:46:04
Job time: 7054 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 10:30:25 ; Search time 1021.9 Seconds
(without alignments)
225.363 Million cell updates/sec

Title: US-09-101-423B-7
Perfect score: 45
Sequence: 1 aatccaagcttgcggccgat.....cgaatatgcggccgcattat 45

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pr1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vi:*
29: gb_ba3:*
30: gb_in1:*
31: gb_in2:*
32: gb_in3:*
33: gb_pl3:*
34: gb_pr4:*
35: em_ba1:*
36: em_ba2:*
37: em_htg1:*
38: em_htg2:*
39: em_htg3:*
40: em_htg4:*
41: em_htg5:*
42: em_htg6:*
43: em_htg7:*
44: em_htg8:*
45: em_htg9:*
46: em_htg10:*
47: em_hum3:*
48: em_hum4:*
49: em_hum5:*
50: em_hum6:*
51: gb_pr5:*
52: gb_pr6:*
53: gb_pr7:*
54: gb_htg1:*
55: gb_htg2:*
56: gb_htg3:*
57: gb_htg4:*
58: gb_htg5:*
59: gb_htg6:*
60: gb_htg7:*
61: gb_htg8:*
62: gb_htg9:*
63: gb_htg10:*
64: gb_htg11:*
65: gb_htg12:*
66: gb_htg13:*
67: gb_htg14:*
68: gb_htg15:*
69: gb_htg16:*
70: gb_htg17:*
71: gb_htg18:*
72: gb_htg19:*
73: gb_htg20:*
74: gb_htg21:*
75: gb_htg22:*
76: gb_htg23:*
77: gb_sts1:*
78: gb_sts2:*
79: gb_vi1:*
80: gb_vi2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		24.8	55.1	86896	2	RCU57682	U57682 Rhodobacter
2	c	23.6	52.4	210047	56	AC010906	AC010906 Homo sapi
3	c	23.4	52.0	177241	52	HS402G11	AL022328 Human DNA
4		23.2	51.6	6254	32	DMU13637	U13637 Drosophila
5		23.2	51.6	14724	58	AC015103	AC015103 Drosophil
6	c	23.2	51.6	218971	56	AC011697	AC011697 Drosophil
7		23.2	51.6	308373	30	AE003495	AE003495 Drosophil
8		23	51.1	10908	1	AE002294	AE002294 Chlamydia
9		22.2	49.3	2013	2	BACCOTJABC	L38014 Bacillus su
10		22.2	49.3	2565	1	AF030293	AF030293 Rathayiba
11		22.2	49.3	2565	2	RRAJ2069	AJ002069 Rathayiba
12		22.2	49.3	10121	1	AE001932	AE001932 Deinococc
13		22.2	49.3	18622	1	AF012532	AF012532 Bacillus
14		22.2	49.3	213190	2	BSUB0004	Z99107 Bacillus su
15		22	48.9	11936	1	AE004219	AE004219 Vibrio ch
16	c	22	48.9	66441	32	PFMAL1P4	AL031747 Plasmodiu
17	c	22	48.9	326301	2	NMA6Z2491	AL162757 Neisseria
18	c	21.8	48.4	11627	1	AE004582	AE004582 Pseudomon
19		21.8	48.4	25458	29	SCE22	AL355832 Streptomy
20		21.6	48.0	47811	59	AC017664	AC017664 Drosophil
21	c	21.6	48.0	119526	54	AC007468	AC007468 Drosophil

	22	21.6	48.0	151673	62	AC022365 AC022365 Homo sapi
	23	21.6	48.0	203634	51	CNS01DVL AL135879 Human chr
	24	21.6	48.0	203650	51	CNS01DSI AL121790 Human chr
	25	21.6	48.0	341319	30	AE003536 AE003536 Drosophil
c	26	21.4	47.6	4566	2	FRANIFX L29299 Frankia aln
c	27	21.4	47.6	5291	2	MLUSTROA M17788 M.luteus st
	28	21.4	47.6	10529	1	AE005138 AE005138 Halobacte
	29	21.4	47.6	11489	1	AE002455 AE002455 Neisseria
c	30	21.4	47.6	35437	29	SCF43A AL096837 Streptomy
c	31	21.4	47.6	42655	29	SC7H2 AL109732 Streptomy
	32	21.2	47.1	611	3	OAINIGFII6 U00667 Ovis aries st
	33	21.2	47.1	1114	67	AC047083 AC047083 Giardia i
c	34	21.2	47.1	3481	2	AX000291 AX000291 Sequence
c	35	21	46.7	331	77	G23787 G23787 human STS W
	36	21	46.7	735	11	AF232928 AF232928 Mus muscu
	37	21	46.7	3711	1	AF202319 AF202319 Rhodospir
c	38	21	46.7	4615	1	AF182519 AF182519 Pseudomon
c	39	21	46.7	5763	33	SCYGR098C Z72883 S.cerevisia
	40	21	46.7	6298	33	YSCESP1A L07289 Saccharomyc
c	41	21	46.7	10186	1	AE004583 AE004583 Pseudomon
c	42	21	46.7	11266	1	AE004461 AE004461 Pseudomon
c	43	21	46.7	103743	54	AC007503 AC007503 Homo sapi
c	44	21	46.7	151802	73	AL354867 AL354867 Homo sapi
c	45	21	46.7	155611	63	AC023826 AC023826 Homo sapi

ALIGNMENTS

RESULT 1
RCU57682
LOCUS RCU57682 86896 bp DNA BCT 07-FEB-1997
DEFINITION Rhodobacter capsulatus cosmids 143-147, complete sequence.
ACCESSION U57682
VERSION U57682.1 GI:1613791
KEYWORDS HTG.
SOURCE Rhodobacter capsulatus.
ORGANISM Rhodobacter capsulatus
Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
Rhodobacter.
REFERENCE 1 (bases 1 to 86896)
AUTHORS Kumar,V., Fonstein,M. and Haselkorn,R.
TITLE Bacterium genome sequence
JOURNAL Nature 381 (6584), 653-654 (1996)
MEDLINE 96260014
REFERENCE 2 (bases 1 to 86896)
AUTHORS Fonstein,M., Kumar,V. and Haselkorn,R.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-1996) M. Fonstein, MGCB, Univ of Chicago, 920 E.
58th, Chicago, IL 60637, USA
FEATURES Location/Qualifiers
source 1. .86896
/organism="Rhodobacter capsulatus"
/strain="SB1003"
/db_xref="taxon:1061"
/map="cosmids 143-147 of the complete encyclopedia"
BASE COUNT 15089 a 28056 c 28787 g 14963 t 1 others
ORIGIN

Query Match 55.1%; Score 24.8; DB 2; Length 86896;
Best Local Similarity 80.6%; Pred. No. 48;
Matches 29; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy       4 ccaagcttgcggccgatcaggccgaatatgcggccg 39
           |||| | ||||||| || ||||||||   |||||||
Db   76652 CCAACCGTGCGGCCTATGAGGCCGAACGCGCGGCCG 76687
```

RESULT 2
AC010906/c
LOCUS AC010906 210047 bp DNA HTG 20-AUG-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-566O4, WORKING DRAFT SEQUENCE,
12 unordered pieces.
ACCESSION AC010906
VERSION AC010906.7 GI:9799871
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 210047)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 210047)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 13, 2000 this sequence version replaced gi:9211497.

---------- Genome Center ----------
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
---------- Project Information ----------
Center project name: H_NH0566O04
---------- Summary Statistics ----------
Sequencing vector: M13; 85%
Sequencing vector: plasmid; 15%
Chemistry: Dye-primer ET; 69% of reads
Chemistry: Dye-terminator Big Dye; 31% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 197418 bases at least Q40
Consensus quality: 202189 bases at least Q30
Consensus quality: 204884 bases at least Q20
Insert size: 203000; agarose-fp
Insert size: 208947; sum-of-contigs
Quality coverage: 4.78 in Q20 bases; agarose-fp
Quality coverage: 4.63 in Q20 bases; sum-of-contigs
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1569: contig of 1569 bp in length
* 1570 1669: gap of unknown length
* 1670 5999: contig of 4330 bp in length
* 6000 6099: gap of unknown length
* 6100 12339: contig of 6240 bp in length
* 12340 12439: gap of unknown length
* 12440 20468: contig of 8029 bp in length
* 20469 20568: gap of unknown length
* 20569 30238: contig of 9670 bp in length
* 30239 30338: gap of unknown length
* 30339 44105: contig of 13767 bp in length
* 44106 44205: gap of unknown length
* 44206 58754: contig of 14549 bp in length
* 58755 58854: gap of unknown length
* 58855 73429: contig of 14575 bp in length
* 73430 73529: gap of unknown length
* 73530 94651: contig of 21122 bp in length
* 94652 94751: gap of unknown length
* 94752 119515: contig of 24764 bp in length
* 119516 119615: gap of unknown length
* 119616 154482: contig of 34867 bp in length
* 154483 154582: gap of unknown length
* 154583 210047: contig of 55465 bp in length.
FEATURES Location/Qualifiers
source 1. .210047

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-566O4"
     misc_feature    1. .1569
                     /note="assembly_name:Contig13"
     misc_feature    1670. .5999
                     /note="assembly_name:Contig14"
     misc_feature    6100. .12339
                     /note="assembly_name:Contig15"
     misc_feature    12440. .20468
                     /note="assembly_name:Contig16"
     misc_feature    20569. .30238
                     /note="assembly_name:Contig17"
     misc_feature    30339. .44105
                     /note="assembly_name:Contig18"
     misc_feature    44206. .58754
                     /note="assembly_name:Contig19"
     misc_feature    58855. .73429
                     /note="assembly_name:Contig20"
     misc_feature    73530. .94651
                     /note="assembly_name:Contig21"
     misc_feature    94752. .119515
                     /note="assembly_name:Contig22
                     clone_end:SP6
                     vector_side:right"
     misc_feature    119616. .154482
                     /note="assembly_name:Contig23"
     misc_feature    154583. .210047
                     /note="assembly_name:Contig24"
BASE COUNT    56553 a  50410 c  48661 g  53321 t   1102 others
ORIGIN


  Query Match             52.4%;  Score 23.6;  DB 56;  Length 210047;
  Best Local Similarity   76.3%;  Pred. No. 1.1e+02;
  Matches   29;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1 aatccaagcttgcggccgatcaggccgaatatgcggcc 38
           ||||||||||| | ||| | ||||| ||| | |  |||
Db 102087 AATCCAAGCTTTCAGCCCAGCAGGCTGAAAAGGATGCC 102050


RESULT   3
HS402G11/c
LOCUS       HS402G11   177241 bp    DNA              PRI       12-DEC-1999
DEFINITION  Human DNA sequence from clone RP3-402G11 on chromosome
            22q13.31-13.33 Contains the MAPK12 gene for mitogen activated
            protein kinase 12 (SAPK3), the MAPK11 gene for mitogen activated
            protein kinase 11 (PRKM11), gene KIAA0315, the gene for a novel
            protein similar to KIAA0901 and mouse histone deacetylase MHDA2,
            the gene for a novel protein similar to Xenopus gamma-tubulin
            interacting protein (yeast SPC98 homolog), the gene for a novel
            protein similar to yeast and bacterial predicted proteins, the gene
            for a novel protein similar to C. elegans F38A5.2, the gene for a
            novel protein similar to MRS1 and the gene for a novel protein
            similar to mouse MOV10 (GB110) and yeast and plant predicted
            proteins. Contains ESTs, GSSs and fifteen putative CpG islands,
            complete sequence.
ACCESSION   AL022328
VERSION     AL022328.21  GI:5263010
KEYWORDS    HTG; CpG island; GB110; KIAA0315; KIAA0901; kinase; MAPK11; MAPK12;
            MHDA2; mitogen activated protein kinase; mitogen-activated; MOV10;
            MRS1; PRKM11; SAPK3; SPC98.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177241)
  AUTHORS   Coville,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 29, 1999 this sequence version replaced gi:5262835.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP3-402G11 is
            from the library RPCI-3 constructed at the Roswell Park Cancer
            Institute by the group of Pieter de Jong. For further details see
            http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP3-402G11 The true
            right end of clone RP5-898I4 is at 23403 in this sequence. The true
            right end of clone RP11-232E17 is at 56734 in this sequence. The
            true right end of clone RP4-600O24 is at 96013 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .177241
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="q13.31-13.33"
                     /clone="RP3-402G11"
                     /clone_lib="RPCI-3"
     repeat_region   135. .447
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   657. .959
                     /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region   1373. .1659
                     /note="AluSx repeat: matches 1. .298 of consensus"
     repeat_region   1669. .1966
                     /note="AluSq repeat: matches 1. .306 of consensus"
     repeat_region   2504. .2574
                     /note="MER58 repeat: matches 170. .243 of consensus"
     repeat_region   2663. .2774
                     /note="L2 repeat: matches 2643. .2747 of consensus"
     repeat_region   2775. .3072
                     /note="AluSp repeat: matches 1. .300 of consensus"
     repeat_region   3073. .3086
                     /note="L2 repeat: matches 2630. .2643 of consensus"
     repeat_region   3959. .4455
                     /note="MER1A repeat: matches 31. .527 of consensus"
     repeat_region   5044. .5270
                     /note="AluJb repeat: matches 77. .303 of consensus"
     repeat_region   6305. .6592
                     /note="AluJo repeat: matches 29. .308 of consensus"
     misc_feature    6406. .6907
                     /note="match: GSS: Em:AQ475194"
     repeat_region   6853. .7155
                     /note="AluY repeat: matches 1. .303 of consensus"
     repeat_region   7218. .7516
                     /note="AluYb8 repeat: matches 1. .318 of consensus"
     repeat_region   9018. .9309
                     /note="AluSx repeat: matches 1. .294 of consensus"
     repeat_region   9427. .9542
                     /note="AluJo repeat: matches 30. .150 of consensus"
     repeat_region   9543. .9847
                     /note="AluY repeat: matches 1. .303 of consensus"
```

```
repeat_region   9848. .10020
                /note="AluJo repeat: matches 150. .306 of consensus"
repeat_region   10024. .10316
                /note="AluSg repeat: matches 1. .293 of consensus"
repeat_region   10324. .10619
                /note="AluY repeat: matches 2. .297 of consensus"
repeat_region   10804. .11101
                /note="AluSg repeat: matches 1. .296 of consensus"
repeat_region   11227. .11641
                /note="MER4C repeat: matches 22. .461 of consensus"
repeat_region   11642. .11957
                /note="AluY repeat: matches 1. .310 of consensus"
repeat_region   11958. .11981
                /note="MER4C repeat: matches 1. .22 of consensus"
repeat_region   14063. .14370
                /note="AluY repeat: matches 2. .310 of consensus"
repeat_region   15488. .15718
                /note="AluSx repeat: matches 51. .295 of consensus"
repeat_region   16411. .16616
                /note="AluY repeat: matches 91. .296 of consensus"
repeat_region   16816. .17110
                /note="AluJb repeat: matches 1. .305 of consensus"
repeat_region   18096. .18300
                /note="MER46C repeat: matches 113. .337 of consensus"
repeat_region   18366. .18558
                /note="MIR repeat: matches 32. .247 of consensus"
repeat_region   18590. .18727
                /note="LTR45 repeat: matches 389. .525 of consensus"
repeat_region   21328. .21889
                /note="MER54B repeat: matches 1. .638 of consensus"
repeat_region   21958. .22115
                /note="L1 repeat: matches 4613. .4770 of consensus"
repeat_region   22116. .22245
                /note="FLAM_C repeat: matches 1. .129 of consensus"
repeat_region   22246. .22887
                /note="L1 repeat: matches 3975. .4613 of consensus"
repeat_region   22938. .23151
                /note="FRAM repeat: matches 1. .166 of consensus"
repeat_region   23152. .23217
                /note="L1 repeat: matches 3919. .3994 of consensus"
repeat_region   23218. .23373
                /note="AluSg/x repeat: matches 133. .302 of consensus"
repeat_region   23375. .23514
                /note="L1 repeat: matches 3780. .3926 of consensus"
repeat_region   23516. .23691
                /note="FRAM repeat: matches -7. .168 of consensus"
repeat_region   23699. .24482
                /note="L1 repeat: matches 2915. .3769 of consensus"
repeat_region   24501. .24747
                /note="L1MEc repeat: matches 1501. .1751 of consensus"
repeat_region   24978. .25194
                /note="L1MEc repeat: matches 272. .492 of consensus"
repeat_region   25933. .26246
                /note="AluY repeat: matches 1. .309 of consensus"
repeat_region   27926. .28225
                /note="AluSx repeat: matches 2. .302 of consensus"
repeat_region   28694. .28997
                /note="AluSg repeat: matches 1. .304 of consensus"
repeat_region   29001. .29119
                /note="AluJo/FLAM repeat: matches 3. .133 of consensus"
repeat_region   29859. .30170
                /note="L1ME repeat: matches 5189. .5489 of consensus"
repeat_region   30171. .30449
                /note="AluSx repeat: matches 24. .304 of consensus"
repeat_region   30450. .30556
                /note="L1ME repeat: matches 5086. .5189 of consensus"
repeat_region   30567. .30813
                /note="L2 repeat: matches 1079. .1303 of consensus"
repeat_region   30975. .31047
                /note="L1MC/D repeat: matches 5325. .5397 of consensus"
repeat_region   31048. .31362
                /note="AluY repeat: matches 1. .303 of consensus"
repeat_region   31363. .31990
                /note="L1MC/D repeat: matches 4762. .5325 of consensus"
repeat_region   31997. .32351
                /note="MLT1A1 repeat: matches 16. .365 of consensus"
repeat_region   32365. .32907
                /note="L1MC/D repeat: matches 4184. .4755 of consensus"
repeat_region   32908. .33208
                /note="AluSx repeat: matches 1. .301 of consensus"
repeat_region   33209. .33464
                /note="L1MC/D repeat: matches 3936. .4184 of consensus"
repeat_region   33468. .33637
                /note="FRAM repeat: matches 1. .175 of consensus"
repeat_region   33650. .33707
                /note="L1MEc repeat: matches 2384. .2075 of consensus"
repeat_region   33708. .33998
                /note="AluSx repeat: matches 1. .288 of consensus"
repeat_region   33999. .34144
                /note="L1MEc repeat: matches 2231. .2385 of consensus"
repeat_region   34781. .34916
                /note="4 copies 34 mer 86 conserved"
repeat_region   34792. .34910
                /note="7 copies 17 mer 73 conserved"
repeat_region   35157. .35466
                /note="AluSg repeat: matches 1. .310 of consensus"
repeat_region   35626. .35922
                /note="AluSg repeat: matches 1. .297 of consensus"
repeat_region   35932. .36042

  Query Match            52.0%;  Score 23.4;  DB 52;  Length 177241;
  Best Local Similarity  73.2%;  Pred. No. 1.4e+02;
  Matches   30;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       4 ccaagcttgcggccgatcaggccgaatatgcggccgcatta 44
             || ||||  | | ||| | ||||||||  | ||||||| ||
Db 101649 CCCAGCTGCCCGGCGAACTGGCCGAATTGGTGGCCGCAGTA 101609


RESULT   4
DMU13637
LOCUS       DMU13637     6254 bp    mRNA            INV       02-MAR-1995
DEFINITION  Drosophila melanogaster vitellogenin receptor (yl) mRNA, complete
            cds.
ACCESSION   U13637
VERSION     U13637.1  GI:535345
KEYWORDS    .
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 6254)
  AUTHORS   Schonbaum,C.P., Lee,S. and Mahowald,A.P.
  TITLE     The Drosophila yolkless gene encodes a vitellogenin receptor
            belonging to the low density lipoprotein receptor superfamily
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 92 (5), 1485-1489 (1995)
  MEDLINE   95183490
REFERENCE   2  (bases 1 to 6254)
  AUTHORS   Schonbaum,C.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-1994) Christopher P. Schonbaum, Molecular
            Genetics and Cell Biology, The University of Chicago, 920 East 58th
            Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .6254
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="gt11 ovary cDNA library of W.R. Steinhauer et
                     al., Mol. Cell Biol. 9:5726(1989)"
                     /map="12E4-12E8; 1-48 cM"
                     /chromosome="1"
                     /sex="female"
                     /tissue_type="ovary"
                     /dev_stage="adult"
     gene            145. .6099
```

```
                /gene="yl"
                /note="yolkless"
                /allele=""
                /db_xref="FlyBase:FBgn0004649"
     CDS        145. .6099
                /gene="yl"
                /standard_name="yolkless"
                /note="putative vitellogenin receptor"
                /codon_start=1
                /protein_id="AAB60217.1"
                /db_xref="GI:535346"
                /translation="MCQAEHQVHPSEQRIRVESPKMTASRRGFNLTSQTRAHPSSGGS
                TSSRYGNCQRTHLIINGRHVAISLLLLVGLCGGTAAGTPGSADTRCDAGQFQCRDGGC
                ILQAKMCDGRGDCKDSSDELDCDYRLCRPPHWFPCAQPHGACLAAELMCNGIDNCPGG
                EDELNCPVRPGFRFGDTAHRMRSCSKYEFMCQQDRTCIPIDFMCDGRPDCTDKSDEVA
                GCKQAEITCPGEGHLCANGRCLRRKQWVCDGVDDCGDGSDERGCLNLCEPQKGKFLCR
                NRETCLTLSEVCDGHSDCSDGSDETDLCHSKPDCDAKKCALGAKCHMMPASGAECFCP
                KGFRLAKFEDKCEDVDECKEQDDLCSQGCENTSGGYRCVCDAGYLLDKDNRTCRAVVY
                GSKEQQPLLLYTTQMTIMGMHLREDNVRNHVYQVAGNLSKVIGVAYDGSHIYWTNIQN
                EAESIVKANGDGSNAEILLTSGLDAPEDLAVDWLTQNIYFSDNIMRHIAVCSNDGLNC
                AVLVTQDVHQPRSLAVWPQKGLMFWTDWGEKPMIGRASMDGSRSRPIVSDNIEWPNGI
                ALDMHQQRIYWVDAKLGSVQTVRPDGTGRRTVLDGMLKHPYGLAIFEDQLYWSDWATK
                SVHACHKFSGKDHRILAKDRTIYAVHIYHPAKQPNSPHGCENATCSHLCLLAEPEIGG
                HSCACPDGMRLAPDHRRCMLMEKRQRLFIGLGQVLLEIEHTAFGAHQVSKSYTLPCLI
                NEMVYNRINGSLIIADNDQRLILEFQPESHESNVLVRSNLGNVSALAFDHLSRNLYWA
                DTERAVIEVLSLQTRHRALIRFFPGQEVPIGLTVMPAEGYLYVVLKAKRHSHIDKIPL
                SGKGEQVHVFEDDLGDDDIKLVTDYETQTIFWSDSDLGRISYSNYRVPHSQIFRGKLR
                RPYSLAMVHHDLFWNELGTPRIYWTHKSNMGPRKVIDIMEKDDPAAIMPYVPVATPNG
                IPLAASSPVGQESHPCQQQNGGCSHICVGEGPYHSICLCPAGFVYRDAGNRTCVEALD
                CEFRCHSGECLTMNHRCNGRRDCVDNSDEMNCDEEHRHKPKVLCSPSQFACHSGEQCV
                DKERRCDNRKDCHDHSDEQHCEKFDKSKKCHVHQHGCDNGKCVDSSLVCDGTNDCGDN
                SDELLCEATLRCEPGMFQCGSGSCIAGSWECDGRIDCSDGSDEHDKCVHRSCPPDMHR
                CLSGQCLDRSLVCDGHNDCGDKSDELNCGTDSSTMNISCAEDQYQCTSNLKICLPSAV
                RCNGTTECPRGEDEADCGDVCSIYEFKCRSGRECIRREFRCDGQKDCGDGSDELSCEL
                EKGHHNQSQIQPWSTSSRSCRPHLFDCQDGECVDLSRVCNNFPDCTNGHDEGPKCATA
                CRSASGRQVCQHKCRATPAGAVCSCFDGYRLDADQKSCLDIDECQEQQPCAQLCENTL
                GGYQCQCHADFMLRQDRVSCKSLQSGATLLFSSFNEVRNLSEQPVMLNVAWSANDSRI
                TGFDLAMHRQMGYFSAEDEGIVYQVDLQTKVIVRALGLPAPTKLSVDWVTGNVYVLSG
                AQEIQACSFVGRMCGRIVHVKSPRHVKHLAVDGYHARIFYIVIRTEGYGQTSSEIHMA
                RLDGSRRDMLLQRSESFMTALTTDPHQQLLYFVDQHMRTLERISYRLKTGPMRRPEIM
                LQKSNALMHPSGLSVYENNAFIVNLGSMEAVQCALYGSRICHKISINVLNAQDIVVAG
                RSRQPQKASHPCAHAHCHGLCLQADYGYECMCGNRLVAEGERCPHGSGNEVAVLGAVN
                SLELEHEHEQNGHFHWLMALFVLAAGSLIAGLGYMYYQYRQRGHTDLNINMHFQNPLA
                TLGGTKAFLEHERAEAGVGFTTETGTVSSRGSNDTFTTTSASSSFAAQQFSVPNALQR
                LLRPRQSASGDPMAQELLLESPSRESKLHALDGGGAGGDGDGGCGVGRQVPDILVADM
                DDDAAKSAGQFGGNYAGNDANARFVS."
     variation  4699
                /gene="yl"
                /note="iso-1 (genomic DNA)"
                /db_xref="FlyBase:FBgn0004649"
                /replace="a"
     variation  4764
                /gene="yl"
                /note="iso-1 (genomic DNA)"
                /db_xref="FlyBase:FBgn0004649"
                /replace="g"
     variation  5106
                /gene="yl"
                /note="iso-1 (genomic DNA)"
                /db_xref="FlyBase:FBgn0004649"
                /replace="c"
     polyA_site 6254
                /note="6 A residues"
BASE COUNT     1394 a   1677 c   1910 g   1273 t
ORIGIN


  Query Match           51.6%;  Score 23.2;  DB 32;  Length 6254;
  Best Local Similarity 70.5%;  Pred. No. 2.6e+02;
  Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      2 atccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
          |||||||| ||| |       || || || ||||||||||| |
```

```
Db   4810 ATCCAAGCGTGCAGTTTCGTGGGACGCATGTGCGGCCGCATTGT 4853


RESULT   5
AC015103
LOCUS       AC015103    14724 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC015103
VERSION     AC015103.1  GI:6436232
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 14724)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10209883 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .14724
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     3605 a   3693 c   3916 g   3510 t
ORIGIN


  Query Match           51.6%;  Score 23.2;  DB 58;  Length 14724;
  Best Local Similarity 70.5%;  Pred. No. 2.3e+02;
  Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      2 atccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
          |||||||| ||| |       || || || ||||||||||| |
Db   8691 ATCCAAGCGTGCAGTTTCGTGGGACGCATGTGCGGCCGCATTGT 8734


RESULT   6
AC011697/c
LOCUS       AC011697   218971 bp    DNA             HTG       15-FEB-2000
DEFINITION  Drosophila melanogaster chromosome X clone BACR12J05 (D884) RPCI-98
            12.J.5 map 12E-12E strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 125 unordered pieces.
ACCESSION   AC011697
VERSION     AC011697.7  GI:6978356
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 218971)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 218971)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
```

Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

TITLE Direct Submission

JOURNAL Submitted (11-OCT-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT On Feb 15, 2000 this sequence version replaced gi:6838484.
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 125 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
*        1      446: contig of 446 bp in length
*      447      526: gap of unknown length
*      527     1055: contig of 529 bp in length
*     1056     1135: gap of unknown length
*     1136     1828: contig of 693 bp in length
*     1829     1908: gap of unknown length
*     1909     2831: contig of 923 bp in length
*     2832     2911: gap of unknown length
*     2912     3946: contig of 1035 bp in length
*     3947     4026: gap of unknown length
*     4027     4843: contig of 817 bp in length
*     4844     4923: gap of unknown length
*     4924     5371: contig of 448 bp in length
*     5372     5451: gap of unknown length
*     5452     5985: contig of 534 bp in length
*     5986     6065: gap of unknown length
*     6066     6592: contig of 527 bp in length
*     6593     6672: gap of unknown length
*     6673     7486: contig of 814 bp in length
*     7487     7566: gap of unknown length
*     7567     7942: contig of 376 bp in length
*     7943     8022: gap of unknown length
*     8023     8473: contig of 451 bp in length
*     8474     8553: gap of unknown length
*     8554     9234: contig of 681 bp in length
*     9235     9314: gap of unknown length
*     9315     9930: contig of 616 bp in length
*     9931    10010: gap of unknown length
*    10011    10680: contig of 670 bp in length
*    10681    10760: gap of unknown length
*    10761    11304: contig of 544 bp in length
*    11305    11384: gap of unknown length
*    11385    12077: contig of 693 bp in length
*    12078    12157: gap of unknown length
*    12158    12741: contig of 584 bp in length
*    12742    12821: gap of unknown length
*    12822    14882: contig of 2061 bp in length
*    14883    14962: gap of unknown length
*    14963    15971: contig of 1009 bp in length
*    15972    16051: gap of unknown length
*    16052    17440: contig of 1389 bp in length
*    17441    17520: gap of unknown length
*    17521    18329: contig of 809 bp in length
*    18330    18409: gap of unknown length
*    18410    19820: contig of 1411 bp in length
*    19821    19900: gap of unknown length
*    19901    21100: contig of 1200 bp in length
*    21101    21180: gap of unknown length
*    21181    22181: contig of 1001 bp in length
*    22182    22261: gap of unknown length
*    22262    23685: contig of 1424 bp in length
*    23686    23765: gap of unknown length
*    23766    25198: contig of 1433 bp in length
*    25199    25278: gap of unknown length
*    25279    27114: contig of 1836 bp in length
*    27115    27194: gap of unknown length
*    27195    28549: contig of 1355 bp in length
*    28550    28629: gap of unknown length
*    28630    30577: contig of 1948 bp in length
*    30578    30657: gap of unknown length
*    30658    32556: contig of 1899 bp in length
*    32557    32636: gap of unknown length
*    32637    33822: contig of 1186 bp in length
*    33823    33902: gap of unknown length
*    33903    36202: contig of 2300 bp in length
*    36203    36282: gap of unknown length
*    36283    37948: contig of 1666 bp in length
*    37949    38028: gap of unknown length
*    38029    40445: contig of 2417 bp in length
*    40446    40525: gap of unknown length
*    40526    42921: contig of 2396 bp in length
*    42922    43001: gap of unknown length
*    43002    45458: contig of 2457 bp in length
*    45459    45538: gap of unknown length
*    45539    47806: contig of 2268 bp in length
*    47807    47886: gap of unknown length
*    47887    50521: contig of 2635 bp in length
*    50522    50601: gap of unknown length
*    50602    54569: contig of 3968 bp in length
*    54570    54649: gap of unknown length
*    54650    56887: contig of 2238 bp in length
*    56888    56967: gap of unknown length
*    56968    59805: contig of 2838 bp in length
*    59806    59885: gap of unknown length
*    59886    63021: contig of 3136 bp in length
*    63022    63101: gap of unknown length
*    63102    66948: contig of 3847 bp in length
*    66949    67028: gap of unknown length
*    67029    69188: contig of 2160 bp in length
*    69189    69268: gap of unknown length
*    69269    74338: contig of 5070 bp in length
*    74339    74418: gap of unknown length
*    74419    80192: contig of 5774 bp in length
*    80193    80272: gap of unknown length
*    80273    85047: contig of 4775 bp in length
*    85048    85127: gap of unknown length
*    85128    91343: contig of 6216 bp in length
*    91344    91423: gap of unknown length
*    91424    97193: contig of 5770 bp in length
*    97194    97273: gap of unknown length
*    97274   110572: contig of 13299 bp in length
*   110573   110652: gap of unknown length
*   110653   124105: contig of 13453 bp in length
*   124106   124185: gap of unknown length
*   124186   135484: contig of 11299 bp in length
*   135485   135564: gap of unknown length
*   135565   155302: contig of 19738 bp in length
*   155303   155382: gap of unknown length
*   155383   171964: contig of 16582 bp in length
*   171965   172044: gap of unknown length
*   172045   172568: contig of 524 bp in length
*   172569   172648: gap of unknown length
*   172649   173373: contig of 725 bp in length
*   173374   173453: gap of unknown length
*   173454   173951: contig of 498 bp in length
*   173952   174031: gap of unknown length
*   174032   174486: contig of 455 bp in length
*   174487   174566: gap of unknown length
*   174567   175636: contig of 1070 bp in length
*   175637   175716: gap of unknown length
*   175717   176341: contig of 625 bp in length
*   176342   176421: gap of unknown length
*   176422   176848: contig of 427 bp in length
```

```
          *   176849   176928: gap of unknown length
          *   176929   177628: contig of 700 bp in length
          *   177629   177708: gap of unknown length
          *   177709   178205: contig of 497 bp in length
          *   178206   178285: gap of unknown length
          *   178286   179653: contig of 1368 bp in length
          *   179654   179733: gap of unknown length
          *   179734   180521: contig of 788 bp in length
          *   180522   180601: gap of unknown length
          *   180602   181263: contig of 662 bp in length
          *   181264   181343: gap of unknown length
          *   181344   181859: contig of 516 bp in length
          *   181860   181939: gap of unknown length
          *   181940   182780: contig of 841 bp in length
          *   182781   182860: gap of unknown length
          *   182861   183108: contig of 248 bp in length
          *   183109   183188: gap of unknown length
          *   183189   183760: contig of 572 bp in length
          *   183761   183840: gap of unknown length
          *   183841   184205: contig of 365 bp in length
          *   184206   184285: gap of unknown length
          *   184286   184909: contig of 624 bp in length
          *   184910   184989: gap of unknown length
          *   184990   185500: contig of 511 bp in length
          *   185501   185580: gap of unknown length
          *   185581   185852: contig of 272 bp in length
          *   185853   185932: gap of unknown length

Query Match              51.6%;  Score 23.2;  DB 56;  Length 218971;
Best Local Similarity    70.5%;  Pred. No. 1.6e+02;
Matches   31;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      2 atccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
          |||||||| ||| |       || || || ||||||||||| |
Db 142092 ATCCAAGCGTGCAGTTTCGTGGGACGCATGTGCGGCCGCATTGT 142049


RESULT   7
AE003495
LOCUS       AE003495   308373 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386053 section 12
            of 30, complete sequence.
ACCESSION   AE003495 AE002593
VERSION     AE003495.2  GI:10728241
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 308373)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
    TITLE   The genome sequence of Drosophila melanogaster
    JOURNAL Science 287 (5461), 2185-2195 (2000)
    MEDLINE 20196006
REFERENCE   2  (bases 1 to 308373)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
    TITLE   Direct Submission
    JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7292953.
FEATURES             Location/Qualifiers
     source          1. .308373
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="X"
     mRNA            join(1418. .1465,1607. .>2164)
                     /gene="CG1368"
                     /product="CT3032"
                     /db_xref="FLYBASE:FBan0001368"
                     /db_xref="FLYBASE:FBgn0030539"
     gene            <1418. .>2164
                     /gene="CG1368"
                     /map="12D2-12D3"
                     /db_xref="FLYBASE:FBan0001368"
                     /db_xref="FLYBASE:FBgn0030539"
     CDS             join(1463. .1465,1607. .2164)
                     /gene="CG1368"
                     /note="CG1368 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0001368"
                     /db_xref="FLYBASE:FBgn0030539"
                     /protein_id="AAF48343.1"
                     /db_xref="GI:7292954"
                     /translation="MKFLIAFALFACVAADVSHLSNEYLPPVQSSYAAPSVSYSAPAV
                     QQTYAAPAIQQSYVAPSNEYLPPVQTYSAPAVQRTYSAPAVQRTYSAPSVSYSAPSVS
                     YSAPSVSYSAPAVQQSYSAPSVSYSAPAVQQSYSAPSVSYSAPAVQQSYSAPAVSYSA
                     PSVSYSAPSVDVGTQYASNGGYVYRK"
     mRNA            <3583. .>4587
                     /gene="CG11581"
                     /product="CT36532"
                     /db_xref="FLYBASE:FBan0011581"
                     /db_xref="FLYBASE:FBgn0030540"
     gene            <3583. .>4587
                     /gene="CG11581"
                     /map="12D2-12D3"
                     /db_xref="FLYBASE:FBan0011581"
                     /db_xref="FLYBASE:FBgn0030540"
     CDS             3583. .4587
                     /gene="CG11581"
                     /note="CG11581 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0011581"
                     /db_xref="FLYBASE:FBgn0030540"
                     /protein_id="AAF48344.1"
                     /db_xref="GI:7292955"
                     /translation="MMQPWAVVVLLLPLINGDVSHLPREYIPPSAELEPTTRSAADGQ
                     EVQKDPSGFYPSTGLPLPPGYAIPTGSNVLPQPVQPPNFPLPIYPPFYGFGGGPVEQG
                     GGHGIGPVPSAYPNGGAFPAGPTNGGAFAGGPTNAAFQLPPGGPFGPQAGNFAQPGGF
```

```
                PQAGGFPPQGAGFQPQGIPGNGFPTQGAPFSPQQGVFPSQGGAFGPQGIPGNGFLGGN
                FPPQGGPFPGSAYPPEFLTNQGLPLTNPQGPFQGPSPAGFNVPAGFGGPVPGQNPYQQ
                FGPGFGGPAPPGYSDEPAKEQPKDSERTVVAPPSADDPKSVADTVYASNGGYVYKRAK
                "
mRNA            join(<7576. .7662,7828. .9753,11338. .>12372)
                /gene="CG11584"
                /product="CT36540"
                /db_xref="FLYBASE:FBan0011584"
                /db_xref="FLYBASE:FBgn0030541"
gene            <7576. .>12372
                /gene="CG11584"
                /map="12D2-12D4"
                /db_xref="FLYBASE:FBan0011584"
                /db_xref="FLYBASE:FBgn0030541"
CDS             join(7576. .7662,7828. .9753,11338. .12372)
                /gene="CG11584"
                /note="CG11584 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0011584"
                /db_xref="FLYBASE:FBgn0030541"
                /protein_id="AAF48345.1"
                /db_xref="GI:7292956"
                /translation="MAYASWYGGISHQLRGELPTINADGASKMKPFAFIVVAALSLVV
                VSGRPGAQYLPPLPVKEVSAAQSFQQLVQTQLQQHIQSQIHPQVESIQASVPLASFTI
                QSSGQQQYQSAPAPAPAPVSIPAPAPVVNSAPAPAPAPVEVQQPAPQVIYEAPKPVIA
                TQTAAKNAYLPPAPVQQFVPAPEPVVANIAPQQETITTTYNAPAAAPVPAPAPIAIPV
                PAVQEQHQVIQQTYSVPAPAPAVQQSYSAPAPAPVVQQTYSAPAPVVQEQTYTAAAPV
                QAVQQLTYSAPAPVTQEQYYSAPASVVQQTSSAPAPAPAPVQEQFYSAPAPAIQQTYS
                APAPAPVVQQTYSAPAPAPQQTYSAPAPAVQEQTQVVQSYSAPAPAPVAQQTYSYPAP
                VVQQAPVVQAVAQQAPVVQQSYSAPAPAPVVQQTYSAPAPVVQETIQQAPVIQQAPVV
                QQSYSAPAPAPVVQQSYSAPAPVVQETIQQAPVIQQAPVVQQSYSAPAPVVQETIQQA
                PVIQQAPVVQQSYSAPAPAPVVQQSYSAPAPAPVVQESIQQAPVIQQSYTAPAPVSIP
                EPVQQIVQQPQYSGYSYQTPQQAPAPIQQSLPAPAPVVISAPAPAPAPAPVQLQQSFV
                PAPPAPIQIQQPAQPIVQYSPPVVQQQVTYTQPAPAPIQVAAAAPVAKFFAILSLALL
                AVASADVSHLTNTYLPPKSAAASTSYESYQAAPVEAAASETYVAPAAAAASTYESESYS
                APAASFTSESYAAPAAVEAAVETAAEETNEQPAASYVAPVVTKTTYSAPAVSSYSSYS
                APAVVAKTYTAPAVVKTYSAPAVSTYSAPAVSGYSQTYTAPAVVKTYSAPAVSTYTAP
                VVTKTYTAPVVAKTYTAPAVSTYTAPVVAKTYTAPAVVKTYSAPAVSTYSAPAVSTYT
                APVVTKTYTAPVVTKTYTAPAVVKTYSAPAVSTYTAPAVSTYTAPVVAKTYSAPAVST
                YTAPVVTKTYSAPAVSSYSAPAVVSSYSGSSGTVYGSNGGYVY"
mRNA            <13948. .>14277
                /gene="CG12481"
                /product="CT32754"
                /db_xref="FLYBASE:FBan0012481"
                /db_xref="FLYBASE:FBgn0030542"
                /evidence=not_experimental
gene            <13948. .>14277
                /gene="CG12481"
                /map="12D3-12D4"
                /db_xref="FLYBASE:FBan0012481"
                /db_xref="FLYBASE:FBgn0030542"
                /evidence=not_experimental
CDS             13948. .14277
                /gene="CG12481"
                /note="CG12481 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0012481"
                /db_xref="FLYBASE:FBgn0030542"
                /evidence=not_experimental
                /protein_id="AAF48346.1"
                /db_xref="GI:7292957"
                /translation="MALLASTGAQPVSQTEMEPMLINVTTSRPKLVIEEMEPMRYHFV
                TEHRPTSLGQNYYIKPGQPVGSITLDSDTLQVRYDQDDGKPVASKHNILFVAYAKDLA
                QKSRQRK"
mRNA            join(<16436. .16455,16506. .>17307)
                /gene="CG11585"
                /product="CT32755"
                /db_xref="FLYBASE:FBan0011585"
                /db_xref="FLYBASE:FBgn0030543"
gene            <16436. .>17307
                /gene="CG11585"
                /map="12D3-12D4"
                /db_xref="FLYBASE:FBan0011585"
                /db_xref="FLYBASE:FBgn0030543"
CDS             join(16436. .16455,16506. .17307)
                /gene="CG11585"
                /note="CG11585 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0011585"
                /db_xref="FLYBASE:FBgn0030543"
                /protein_id="AAF48347.1"
                /db_xref="GI:7292958"
                /translation="MPAGPTPYSSGGSSYAAPSVSHVSYSAPAASASASHVSYSAPAA
                AHSSYAAPSVSHGSYSSASRTSYSAPVAAPSRKYLAPAAVSHSSYSAPVVSRKSYSAP
                AVSHGSYSSSGGSSGSYSAPVASYSSYSAPAASHTSYSAPVAAPSRQYLAPAASSYSA
                PAVSSYSAPAVSSYSAPAVSSYSAPAVSHGSSYSTSSLSHGSYAAPSVSSKTYAAPAV
                SHGSYSGSSSGSGHGYGSGSSHGSGVRSGHGSSFGSGHKSGSYAANGGYEYRLKH"
mRNA            complement(join(<17993. .18361,18516. .>18623))
                /gene="CG13403"
                /product="CT32756"
                /db_xref="FLYBASE:FBan0013403"
                /db_xref="FLYBASE:FBgn0030544"
                /evidence=not_experimental
```

Query Match 51.6%; Score 23.2; DB 30; Length 308373;
Best Local Similarity 70.5%; Pred. No. 1.5e+02;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```
Qy       2 atccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
           |||||||| ||| |       || || || ||||||||||| |
Db   25075 ATCCAAGCGTGCAGTTTCGTGGGACGCATGTGCGGCCGCATTGT 25118
```

```
RESULT   8
AE002294
LOCUS       AE002294    10908 bp    DNA             BCT       26-MAY-2000
DEFINITION  Chlamydia muridarum, section 26 of 85 of the complete genome.
ACCESSION   AE002294 AE002160
VERSION     AE002294.1  GI:7190305
KEYWORDS    .
SOURCE      Chlamydia muridarum.
  ORGANISM  Chlamydia muridarum
            Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE   1  (bases 1 to 10908)
  AUTHORS   Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
            White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
            Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
            Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
            McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
  TITLE     Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
            pneumoniae AR39
  JOURNAL   Nucleic Acids Res. 28 (6), 1397-1406 (2000)
  MEDLINE   20150255
   PUBMED   10684935
REFERENCE   2  (bases 1 to 10908)
  AUTHORS   Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
            White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
            Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
            Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
            McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10908
                     /organism="Chlamydia muridarum"
                     /db_xref="taxon:83560"
                     /note="synonym: Chlamydia trachomatis MoPn"
     gene            241. .495
                     /gene="TC0265"
     CDS             241. .495
                     /gene="TC0265"
                     /note="hypothetical protein; identified by Glimmer2;
                     putative"
                     /codon_start=1
                     /transl_table=11
```

/product="hypothetical protein"
/protein_id="AAF39134.1"
/db_xref="GI:7190306"
/translation="MIIVGALRDPSWLYYFFLDPFTHVFEKSLSKTRNNKKGYSLFQI
FWGHQKSCFLFHKVKVSTWFVPYKVVGFFVQRLLDHLLKK"

gene 426. .1016
/gene="TC0266"

CDS 426. .1016
/gene="TC0266"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39135.1"
/db_xref="GI:7190307"
/translation="MVCPLQGCGLLRSEASGSSSEEVNQDPLLDAISYSVRVCLAVAK
HTESAKIYHQNRKRCHKVGCWGKVKRFLCCCCCAEKEWVEGKITGLGLWMQEHIRLHG
ILITGYAVHLSGVCWENMLLAGETLSPGDQRQLNAALENSSITWMRLMLDSSDGKVDG
LFSPAMQEDLFAQYREKVGIGNLADAVALRHFCNRT"

gene complement(1104. .3707)
/gene="TC0267"

CDS complement(1104. .3707)
/gene="TC0267"
/note="similar to GP:4376270; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="polymorphic membrane protein G family"
/protein_id="AAF39136.1"
/db_xref="GI:7190308"
/translation="MRPDHVNLCCLCATILSPTAILFGQDALDKSALITKNPNSIVCT
FLEDCTMENFSPALLSHARQDDPLYIIGNTHNWFVSNLHPSTNEERFLKEKGDLSIQD
FRFLSFTDCSSSTEDSPSILYHKNGQLFLRNNGNMSFYRNHSEGSGGALSTDALFLQH
NYLFTNFEENSSAKNGGAIQAQTLSLSRNVSSLSFSRNRANLNGGAICCQNLICSGNV
NPLFFTNNSALNGGAICCINEQNLSEKGCLSLAYNQETLFSGNSAKEKGGAIYTKHMV
LRHNGPVSFVNNSAKLGGAIAIQSGGSLSIIAGGGSVLFQNNSCHFSDQGTVRNAIYL
EKNALLSSLEARHGDILFFDPIVQEVVSPEFSTTSALTPLRIQTNTNRAVIFSSENLS
KEEKTEANLISKIQQPIELQSGCLVLKDRVILSAPSLSQAPQALLVMDVGTSLTTSSD
LKLTTLSIPLHSIDTENSVSIQSPTLSIQKIFLSNSEHENFYENVELLSKDQKDIPLL
SLPKGLPHPDLPDGNLSSHFGYQGDWNFSWQTSDQRETLVANWTANSYIPHPERQSAL
VANTLWNTYSDMQAVQSMINTTAQGGAYLFGTWGSAVSNLFYSHGNSGKSTDNWKHRS
LGYLFGISTHSLDDHSFCLAAGQLFGKSSDSFVTSADTTSYIAAIQTQIATSLIKISA
QACYNESIHELKTKYRSFSKEGFGAWHSVAVSGEIGASIPIVSNGSGLFSSFSIFSKL
QGFSGKQDGFEESRGEARAFADSSFTNISLPVGIAFEKKSQKTRNYYHFLGAYIQDLK
RCVESGPVTLLKNSVTWDAPMANLDSRAWMFRLTNQRALHRFQTLVNMSYMLRGQSYS
YSLDLGTTYRF"

gene 3956. .5800
/gene="TC0268"

CDS 3956. .5800
/gene="TC0268"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39137.1"
/db_xref="GI:7190309"
/translation="MFYFLGWFVMGIKGVGGSGHSDYPIPSHNGDGESEKNSSDSTSS
KVNAKVTSSLQGAPSTNDENSVSPYSVVDVTDLIESGESSRHVIKKSIETEEAAHRES
SVEGAGHSSRGIFGRLQAGLGRLARRVGEAVRNTVGSIFPQRAGAEQRTGKARTKYSP
SASRGLRLMFTDFWRYRVLHRNPPMDGLFAKLDADEAEDMAAYTKEYVSNLEKRGAAD
RETIEHCQMVAKNWEKRARDLRDMGAAKKFLRDPFGKSDPKYKGTLPGEYTVGNTMFY
DGPGVSKLSEVDTGFWLDMEKLSDAVLSANIQKGLRARFVLNQSIPQLESLEERFRKL
ESACDEARASLKEAGWIKEGKEPNKAQRAFRRFVEESRNLELSFGSFGESARRLSARV
SQGLAAAGEAIRRCFDCRKGKYSLKKDLSSEELNLAEELIRFTDEMGIERDPDGNYNI
PWVENWRTGVPVIEGEGAEHIYETMMPVQESFEQVYEVMDMGLEERRDFAVSQQHYQV
PPRSSLNYETPRFREYDVPRNSARSYYDVPRVPPQNEVEEMHVTKGMRSSVYACFVAG
MRNYIVSQPQEQIPNSEQVEQLFQELINDGDQIIQELMKIWNEELDNQ"

gene complement(5923. .6192)
/gene="TC0269"

CDS complement(5923. .6192)
/gene="TC0269"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39138.1"
/db_xref="GI:7190310"
/translation="MLCKVCRGLSSLIVVLGAINAGILGATGYRVNIVARLLGEGTML
TQAAYVVAGIAGVIAFLNFFKCCFKKRHSDCCSSKGGCHHHTDRE"

gene 6393. .6695
/gene="TC0270"

CDS 6393. .6695
/gene="TC0270"
/note="glutamyl-tRNA(Gln) amidotransferase subunit C,
putative; identified by match to TIGR protein family HMM
TIGR00135"
/codon_start=1
/transl_table=11
/product="glutamyl-tRNA(Gln) amidotransferase subunit C,
putative"
/protein_id="AAF73542.1"
/db_xref="GI:8163182"
/translation="MTQSYVNKEEIISLAKNAALELEEAHVEEFVTSMNNVIALMQEV
VAIDISDIILEATVHHFVGPEDLREDMVTSDFTREEFLANVPVSLGGLVKVPTVIK"

gene 6707. .8182
/gene="TC0271"

CDS 6707. .8182
/gene="TC0271"
/note="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamyl-tRNA(Gln) amidotransferase subunit A"
/protein_id="AAF39139.1"
/db_xref="GI:7190311"
/translation="MYRKSALELRDAVVNREISVTAITEYFYHRIESYDEQIGSFLSL
CKERALLRASRIDDKLAKGDPIGILAGIPIGVKDNIHITGVKTTCASKMLENFVAPFD
ATVVRRIEMEDGILLGKLNMDEFAMGSTTRYSAFQHTNNPWDLERVPGGSSGGSAAAV
SARFCPIALGSDTGGSIRQPAAFCGVVGFKPSYGAVSRYGLVAFGSSLDQIGPLTTVV
EDVALAMDAFAGRDIKDATTRDFFRGTFSQALSLEVPKLIGVPRGFLDGLQEDCKENF
FEALAVMERQGSRIIDIDLSVLKHAVPVYYIVASAEAATNLARFDGVRYGHRCAQADN
MQEMYARSRKEGFGKEVTRRILLGNYVLSAERQNIFYKKGTAVRATLIEAFQSAFECC
DVIAMPVCASPAIRDTDVLDPVSLYLQDIYTVAVNLAYLPAISVPSGLSKEGLPLGVQ
FIGKRGADQQICQVGYSFQEHSQIKQLYPKAVNGLFDGGME"

gene 8184. .9650
/gene="TC0272"

CDS 8184. .9650
/gene="TC0272"
/note="similar to PID:1653395; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamyl-tRNA(Gln) amidotransferase subunit B"
/protein_id="AAF39140.1"
/db_xref="GI:7190312"
/translation="MGIAHTEWESVIGLEVHVELNTASKLFSPARNHFGDEPNTNISP
VCTGMPGALPVLNKDAVRKAVLFGCAVEGDVALFSRFDRKSYFYPDSPRNFQITQYEH
PIVRGGCVRAIVEGEEKTFELAQTHLEDDAGMLKHFGDFAGVDYNRAGVPLIEIVSKP
CMFSAEDAVAYANALVSILSYIGISDCNMEEGSVRFDVNISVRPKGSKELRNKVEIKN
MNSFTFMAQALEAEKHRQIEEYLSHPNEDPKKVVPAATYRWDPEKKKTVLMRLKERAE
DYMYFVEPDLPVLQITESYIDEVRQTLPELPHSKYMRYITDFDLAEDLAMILVSDRHT
AHFFETATMSCKNYRALSNWITVEFAGRCKATGKTLPFTGILPEWVAQLVNFIDRGVI
TGKIAKEIADKMVSSFGESPEDILRRHPSLLPMTDDHALRAIVKEVVAQNAASVEDYK
NGKAKALGFLVGQIMKRTEGKAPPKRVNELLLAAMRDA"

gene complement(9731. .10834)
/gene="TC0273"

CDS complement(9731. .10834)
/gene="TC0273"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"

```
                     /protein_id="AAF39141.1"
                     /db_xref="GI:7190313"
                     /translation="MVKIMAPITPTTSPQVKGLLSRFLTAPDRHPKLRYVYDISLIAI
                     SILCIVSIILWTQGSGLALFAIAPALAIGALGVTLLVSDLAESPKSKEVADTVAAVSL
                     PFILTGTAAGLMFSAIAVGGGAVILANPLFLMGSMTLGFALMSLHKVTYQYLSNRSQW
                     QKQNKIKQIESAAWENKLPKESKESSLQTSVRYSSLARKDKTRRNKPGMPNKGSQVPA
                     SIANTERSLRSEEVLHSQSLLRQKELFPNTSNIKKELPNTKSILHTPLNRRSPSGSDS
                     DDVYYTPRAGLSSAETSALGDISGISSSSTSSKTSTPKAKRRVVRSSRSERNARHHRN
                     KEDHRQNQEESSDDEDSSPLPSPRRKKYRSRPK"
BASE COUNT      3210 a    1860 c    2587 g    3251 t
ORIGIN


  Query Match             51.1%;  Score 23;  DB 1;  Length 10908;
  Best Local Similarity   74.4%;  Pred. No. 2.8e+02;
  Matches   29;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       4 ccaagcttgcggccgatcaggccgaatatgcggccgcat 42
           | |||||||  |||||| |||| ||| ||  ||| || |
Db    4524 CAAAGCTTGATGCCGATGAGGCTGAAGATATGGCAGCTT 4562


RESULT   9
BACCOTJABC
LOCUS       BACCOTJABC    2013 bp     DNA              BCT       15-JUN-1996
DEFINITION  Bacillus subtilis cotJABC operon, polypeptide(s) affecting spore
            coat composition, cds.
ACCESSION   L38014
VERSION     L38014.1  GI:1377748
KEYWORDS    .
SOURCE      Bacillus subtilis (strain 168) DNA.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 2013)
  AUTHORS   Henriques,A.O., Beall,B.W., Roland,K. and Moran,C.P. Jr.
  TITLE     Characterization of cotJ, a sigma E-controlled operon affecting the
            polypeptide composition of the coat of Bacillus subtilis spores
  JOURNAL   J. Bacteriol. 177 (12), 3394-3406 (1995)
  MEDLINE   95286532
FEATURES             Location/Qualifiers
     source          1. .2013
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            326. .574
                     /gene="cotJA"
     CDS             326. .574
                     /gene="cotJA"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB17588.1"
                     /db_xref="GI:1377749"
                     /translation="MKDMQPFTPVKSYTPFHSRFDPCPPIGKKYYRTPPNLYMTFQPE
                     HMEQFSPMEALRKGTLWKDLYDFYENPYRGGDAHGKKG"
     gene            519. .821
                     /gene="cotJB"
     CDS             519. .821
                     /gene="cotJB"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB17589.1"
                     /db_xref="GI:1377750"
                     /translation="MIFMKTLIEGETHMAKKVDAEYYRQLEQIQAADFVLVELSLYLN
                     THPHDEDALKQFNQYSGYSRHLKRQFESSYGPLLQFGNSPAGKDWDWGKGPWPWQV"
     gene            836. .1405
                     /gene="cotJC"
     CDS             836. .1405
                     /gene="cotJC"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB17590.1"
                     /db_xref="GI:1377751"
                     /translation="MWVYEKKLQYPVKVSTCNPTLAKYLIEQYGGADGELAAALRYLN
                     QRYTIPDKVIGLLTDIGTEEFAHLEMIATMVYKLTKDATPEQLREAGLGDHYVNHDSA
                     LFYHNAAGVPFTASYIQAKGDPIADLYEDIAAEEKARATYQWLIDISDDPDLNDSLRF
                     LREREIVHSMRFREAVEILKEERDKKKIF"
BASE COUNT      515 a    426 c    553 g    519 t
ORIGIN


  Query Match             49.3%;  Score 22.2;  DB 2;  Length 2013;
  Best Local Similarity   69.8%;  Pred. No. 6.8e+02;
  Matches   30;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       3 tccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
           ||  || ||||||| | ||| || | ||||  | || |||| |
Db     146 TCAGAGTTTGCGGCTGCTCATGCTGCATATTGGACCTCATTTT 188


RESULT  10
AF030293
LOCUS       AF030293     2565 bp    DNA              BCT       13-NOV-1997
DEFINITION  Rathayibacter rathayi protein kinase KdpD gene, complete cds.
ACCESSION   AF030293
VERSION     AF030293.1  GI:2613048
KEYWORDS    .
SOURCE      Rathayibacter rathayi.
  ORGANISM  Rathayibacter rathayi
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Micrococcineae; Microbacteriaceae; Rathayibacter.
REFERENCE   1  (bases 1 to 2565)
  AUTHORS   Labadie,J.C.
  TITLE     Cloning and sequencing of a kinase gene from the bacteria
            Rathayibacter rathayi equivalent to the kdpD gene from Escherichia
            coli
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 2565)
  AUTHORS   Labadie,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-1997) TPA, INRA, St Genes Champanelle 63122,
            France
FEATURES             Location/Qualifiers
     source          1. .2565
                     /organism="Rathayibacter rathayi"
                     /strain="2J"
                     /db_xref="taxon:33887"
     CDS             1. .2565
                     /note="similar to E. coli sensor protein KdpD"
                     /codon_start=1
                     /transl_table=11
                     /product="protein kinase KdpD"
                     /protein_id="AAB84261.1"
                     /db_xref="GI:2613049"
                     /translation="MLVGARVLLGAPRGVGKTFTMLEEAKPLRDEGVDVVVAVVETHG
                     RAGTAAALVGLEVVPRLTVEHRGVVLTEMDVAGVIVRRAPQLALVDELAHTNASRQRT
                     EKRWQDVEAILDAGIDVMSTVNIQHIESLTDVVHKITGAPQRETIPDEVLRAAREIEV
                     IDVTPVAAGALASGLVYPAERIDAALSNYFRLGNLTGLRELALLWLADEVDSALKNYR
                     AEQGIDSTWETRERVVVALTGGPEGDSHPPGATHCRPCGRGELLAVHVTGQDGFAPPT
                     RGLWRQRSLVESLGGSYHQVIGDDIALVEFARAANATQLVIGVSRRGRLARRCPVRGS
                     VDGHRESGNIDVHIVNHAAAGGRFTLPRMAGGALTVKRRLSGLALTLILGPLITAVLV
                     TFRSPDSITSDVLTYQVLVVLVALVGGIWPALLAAVLSGITLDYFLVEPLFTVTVDKP
                     LHLFALALYITIAMMVSYVVDQAARRTRVARRSAAESELLATIAGSVLRGDGALQSLV
                     SRTRKVWVEGCDCWMPRVRPITPTRPPGADGQTAADHAVICADGEPASDDRVVLVPVG
                     ERATLELHGADLDASERRLLAVIAAQIDAALEHEALSVTAREVGPLAETDRVRTALLS
                     AVSHDLRRPFDGGNQKGGWLALHRDDPVCRRPGGAARDRRRKPAHLSVLVTDLLDVSR
                     VQAGVLGVTVRQVDVEDVLPRALDELGVGPDQVVLDLDAAVGPVLADPGLLQRVLVNL
                     LANALRFSPEGAVPTIDQSFGDTVQIRVTDHGPGIAADRRDDVFVPFQRLGDTDNSTG
                     LGLGLALSKGFTVGMGGELDTEDTPGGGLTMVVTLPVASADADANADRPGGSRASL"
BASE COUNT      388 a    771 c    931 g    475 t
ORIGIN
```

```
  Query Match             49.3%;  Score 22.2;  DB 1;  Length 2565;
  Best Local Similarity   77.1%;  Pred. No. 6.6e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       4 ccaagcttgcggccgatcaggccgaatatgcggcc 38
           ||| || |||||||||||| |||||  | ||  ||
Db    1570 CCACGCGTGCGGCCGATCACGCCGACAAGGCCTCC 1604


RESULT  11
RRAJ2069
LOCUS       RRAJ2069     2565 bp    DNA             BCT       24-OCT-1997
DEFINITION  Rathayibacter rathayi kdpD gene, complete.
ACCESSION   AJ002069
VERSION     AJ002069.1  GI:2564304
KEYWORDS    kdpD gene.
SOURCE      Rathayibacter rathayi.
  ORGANISM  Rathayibacter rathayi
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Micrococcineae; Microbacteriaceae; Rathayibacter.
REFERENCE   1  (bases 1 to 2565)
  AUTHORS   Labadie,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-1997) Labadie J.C., TPA, INRA, INRA Centre de
            Theix, 63122 St Genes Champanelle, FRANCE
REFERENCE   2  (bases 1 to 2565)
  AUTHORS   Labadie,J.C.
  TITLE     Clonage et sequencage d'un gene d'une kinase de la bacterie
            Rathayibacter rathayi equivalent au gene KdpD de Escherichia coli
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .2565
                     /organism="Rathayibacter rathayi"
                     /strain="2 J"
                     /db_xref="taxon:33887"
     gene            1. .2565
                     /gene="kdpD"
     CDS             1. .2565
                     /gene="kdpD"
                     /codon_start=1
                     /transl_table=11
                     /product="protein kinase"
                     /protein_id="CAA05169.1"
                     /db_xref="GI:2564305"
                     /db_xref="SWISS-PROT:O34971"
                     /translation="MLVGARVLLGAPRGVGKTFTMLEEAKPLRDEGVDVVVAVVETHG
                     RAGTAAALVGLEVVPRLTVEHRGVVLTEMDVAGVIVRRAPQLALVDELAHTNASRQRT
                     EKRWQDVEAILDAGIDVMSTVNIQHIESLTDVVHKITGAPQRETIPDEVLRAAREIEV
                     IDVTPVAAGALASGLVYPAERIDAALSNYFRLGNLTGLRELALLWLADEVDSALKNYR
                     AEQGIDSTWETRERVVVALTGGPEGDSHPPGATHCRPCGRGELLAVHVTGQDGFAPPT
                     RGLWRQRSLVESLGGSYHQVIGDDIALVEFARAANATQLVIGVSRRGRLARRCPVRGS
                     VDGHRESGNIDVHIVNHAAAGGRFTLPRMAGGALTVKRRLSGLALTLILGPLITAVLV
                     TFRSPDSITSDVLTYQVLVVLVALVGGIWPALLAAVLSGITLDYFLVEPLFTVTVDKP
                     LHLFALALYITIAMMVSYVVDQAARRTRVARRSAAESELLATIAGSVLRGDGALQSLV
                     SRTRKVWVEGCDCWMPRVRPITPTRPPGADGQTAADHAVICADGEPASDDRVVLVPVG
                     ERATLELHGADLDASERRLLAVIAAQIDAALEHEALSVTAREVGPLAETDRVRTALLS
                     AVSHDLRRPFDGGNQKGGWLALHRDDPVCRRPGGAARDRRRKPAHLSVLVTDLLDVSR
                     VQAGVLGVTVRQVDVEDVLPRALDELGVGPDQVVLDLDAAVGPVLADPGLLQRVLVNL
                     LANALRFSPEGAVPTIDQSFGDTVQIRVTDHGPGIAADRRDDVFVPFQRLGDTDNSTG
                     LGLGLALSKGFTVGMGGELDTEDTPGGGLTMVVTLPVASADADANADRPGGSRASL"
BASE COUNT      388 a    771 c    931 g    475 t
ORIGIN


  Query Match             49.3%;  Score 22.2;  DB 2;  Length 2565;
  Best Local Similarity   77.1%;  Pred. No. 6.6e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       4 ccaagcttgcggccgatcaggccgaatatgcggcc 38
           ||| || |||||||||||| |||||  | ||  ||
Db    1570 CCACGCGTGCGGCCGATCACGCCGACAAGGCCTCC 1604


RESULT  12
AE001932
LOCUS       AE001932    10121 bp    DNA             BCT       22-NOV-1999
DEFINITION  Deinococcus radiodurans R1 section 69 of 229 of the complete
            chromosome 1.
ACCESSION   AE001932 AE000513
VERSION     AE001932.1  GI:6458481
KEYWORDS    .
SOURCE      Deinococcus radiodurans.
  ORGANISM  Deinococcus radiodurans
            Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
REFERENCE   1  (bases 1 to 10121)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Fraser,C.M. et,al.
  TITLE     Genome sequence of the radioresistant bacterium Deinococcus
            radiodurans R1
  JOURNAL   Science 286 (5444), 1571-1577 (1999)
  MEDLINE   20036896
REFERENCE   2  (bases 1 to 10121)
  AUTHORS   White,O., Eisen,J.A., Heidelberg,J.F., Hickey,E.K., Peterson,J.D.,
            Dodson,R.J., Haft,D.H., Gwinn,M.L., Nelson,W.C., Richardson,D.L.,
            Moffat,K.S., Qin,H., Jiang,L., Pamphile,W., Crosby,M., Shen,M.,
            Vamathevan,J.J., Lam,P., McDonald,L., Utterback,T., Zalewski,C.,
            Makarova,K.S., Aravind,L., Daly,M.J., Minton,K.W.,
            Fleischmann,R.D., Ketchum,K.A., Nelson,K.E., Salzberg,S.,
            Smith,H.O., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .10121
                     /organism="Deinococcus radiodurans"
                     /strain="R1"
                     /db_xref="taxon:1299"
                     /chromosome="1"
     gene            172. .1347
                     /gene="DR0770"
     CDS             172. .1347
                     /gene="DR0770"
                     /note="similar to GB:U12892 PID:530856 percent identity:
                     53.56; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="fimbrial assembly protein PilM, putative"
                     /protein_id="AAF10349.1"
                     /db_xref="GI:6458482"
                     /translation="MTSLINRFRSRSAAIGVEIGTSTIKVVALKAGAPPSLQHAVMVP
                     TPIGSMRDGLVVEPQAVASELKSLLAEHRITTRHAVTAVPNQVAVTRNIMVPRMDRKD
                     LQSAIRWEAERYIPYPIDEVTLDFDLLDDPANVPEDGQMEVVIAAAPTEAVHRQIEVL
                     RLAGLEPTVVDLKSFAALRALRGNLLGEHLTKSTLTGTNYTEAGEVALVMEIGASSSV
                     INLVRGDRILMTRNINVSADDFTTALQKAFDLDFAAAEDVKLGYATATTPTEDEEDLL
                     NFDLSREQYSPARVFEVVRPVLGDLITEIRRSLEFYRVQSGDVVIDRTFLAGGGAKLR
                     GLAAAIGDALGFGVEVASPWLTVQTDQAGVDTGYLQTNAPEFTVPLGLALRGVMGRG"
     gene            1340. .2056
                     /gene="DR0771"
     CDS             1340. .2056
                     /gene="DR0771"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF10354.1"
                     /db_xref="GI:6458487"
                     /translation="MVEINLLPKEYRRQSQPSVWKYASWAAAGLTAAVLGGWFLSVSG
                     DTNQLRERSAALQQQIDAVAPQKSRFNDLTAQQGELERVTQVATQLRDQKTYWSNDLA
                     SFVERVPGNVVFSNVNMSTVAPGAEPSLAYAGKPVTRQLDLTGSARSQEAIVGFLNAF
                     ETDSNFGVDFKGMQHDATNGIYTFTASIGVVGDQPSAAPGADPTAQGAPATGTAPAAP
                     AAPAPTATTPAAAPAQGGAQ"
     gene            2053. .2712
                     /gene="DR0772"
```

```
CDS             2053. .2712
                /gene="DR0772"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF10355.1"
                /db_xref="GI:6458488"
                /translation="MKAIKLAPHYIFALVLTLCLVLGYLFYTMAIQPRQLEITSLNDE
                ITNKETTLAADQAKAARVPTLTAEVARLEVEREKFLRALPPTANFGQVVANLRQTVSA
                AGGDLKTLNFAGSGAAGANLPAGVRPIGMTMSVNGRFPQLFQILRSLELQNRFTTVDN
                VSLQTQGDAGTGGGGTLGSTLGLTVYTFDASGATSTPDAAAPAAGTPAPAAPAAGGTQ
                P"
gene            2709. .4616
                /gene="DR0773"
CDS             2709. .4616
                /gene="DR0773"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAF10356.1"
                /db_xref="GI:6458489"
                /translation="MSRFPTEDQTQTQTMTEVTNVSQEKSPLNLNFSREVKLVLVVLG
                MVALIGGWFVLTGQPDDQTVTAPVPVTTPAPETATTPETAPAGNAAPDGQTAGTATET
                TTTPGTTDESSGKKATGTAAKSQKGTQSGDGSDESLVATVPPLEGEGTATPPVPVPGG
                INPDRPLKTLSGADPFGSLTPSPGTGTAGGTAPATPVEAQTPETQSPETAPETQAPTA
                QSTTNSPAPVVLSSPAGSSVSGQGGVLDYSSRLDQGALPTPVIPVSPTEVRSPVTSAP
                VSGSSVNGDSVGSGMATAPTVVVSTPAQPTTTPSASSPVARNTAGQSGAGQNTGASRP
                ASGTPSRASAPNVVVVRPPARVSGTSAAGATGSSRTGAAGTPAATPTRNQSASGTRPA
                SGQTPPRVIAVAPAGGARTPAAPGNRPATAAPTRPSPSSTAPSKTVVVRPPVPVSGQQ
                PPRVVTSPAAGVPAGNAAPAPTPPPVVAARVPQANTVPELVAQAGTSAAPATSGASAG
                TGTAGQAGNDAAGTAALPPLITAVGSEAAPASTATADAGSSTPAAPAATALDRLLAPG
                QVALSSVVLGPTNTAVFRTPQGFVVVEQGQTVPGLTGEDGTPVVLQKVETGAVTVALG
                TKVKSLELEQR"
gene            4620. .6842
                /gene="DR0774"
CDS             4620. .6842
                /gene="DR0774"
                /note="similar to SP:Q01565 SP:P31700 GB:X65265 PID:581156
                percent identity: 48.99; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="general secretion pathway protein D, putative"
                /protein_id="AAF10350.1"
                /db_xref="GI:6458483"
                /translation="MNKRHALLLTAVLGMATAYAQTAPTTTTVNTLQTVYRDPSLTSA
                PITANVGKYVGPLSTFLASIAKSAGYEVVFNFNIDALALINGEIVFGNSTASVTTSYA
                TPLGRPQELPAKPVVHNFSNAPFNEAWPLLMDVYELDYQLVKVGSANVIRIGQRPKQL
                ALPLKFISAESALTAIEKFFGEEKFETVISLDSNNKPFQTTRPTGKFGLPNSIKVIPD
                SSNKRLIIGSNSEDGIRIRSFVETIDVQSSGKVISTDSISEIYIVRGQKESVLQFLRD
                SFPELIVTDYASGGLAIEGPRTSVNRAIILLGQVDRAPEIPIVQRIYTVRGQAADITA
                LLAAQYPTLRVTPVGQTGQLVLNGAQAQLDTALALLEQVDRPAPVAESRTVQRVFQLV
                NASAEEVKATLEGTLARDLTADSNNDVLPNVPVTATDANGNTTVVSVPNALGKTANQG
                TANAQAQTAQTPANTQQATLIADKRTNSLIVRGTPEQVAQVAELVPQLDQVVPQINVQ
                VRIQEVNERALQSLGLNWRATFGGFNVAVSGGTGLAATFNPTQSFLGFNIFPTLTALE
                TQGLTRRVYDGNVTMQSGQRSLSATGGAQNASSGAAASVKSGGRLEINIPSAAGNIVR
                QIDYGLNLDFFSPQVAPDGTITLRIRGQVNQPATAITADSLPNLIDFTNSEAQSTITF
                KNGQTILMSGLLGSTETTNRSGVPFLSSLPGVGAAFGEKRTEKTQSQLLVIITGTVVK
                "
gene            6938. .8254
                /gene="DR0775"
CDS             6938. .8254
                /gene="DR0775"
                /note="similar to GB:AE000657 percent identity: 68.57;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="chorismate synthase"
                /protein_id="AAF10351.1"
                /db_xref="GI:6458484"
                /translation="MFVLALLAKKLCHLFVKCSTSYRRLSAKGRRFFDSRGGACLLKP
                PRRGRQSRYSRPMRYLTAGESHGPQLTAIIEGLPAQLPLGKADIDPWLRKRQGGYGRG
                RRMVIETDEAELLSGVRAGRTTGAPVTLAIQNKDHRNWTEIMSPEPGGEPRKKALTDA
                RPGHADLTGGIKYRHKDLRDVLERASARETAARVAVGSIALKLLSELGIEGANYVFNL
                AGIETRQAFSWDALDAIEDSDLRTPDADAAAQMRERIDQAKKDGDTLGGILEVRFRGL
                PVGLGSFVHYDRKLDGKIAQACLSVQAMKGVEIGRAFENAVKPGSGVHDAIHYREGTY
                ARDTNGAGGLEAGMTNGEELIVRVAMKPIATLMKPLPTVNVVTHEASDAARERSDTTA
                VPAAGVILQCVIGWVLAEAITEKFGGDTLPELQERLAAARRYAAEY"
gene            8262. .8978
                /gene="DR0776"
CDS             8262. .8978
                /gene="DR0776"
                /note="similar to GB:AE000657 percent identity: 60.25;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="shikimate kinase"
                /protein_id="AAF10352.1"
                /db_xref="GI:6458485"
                /translation="MTFSELSESVQHALDEVLAAPPRPPEGEAGTPLQGDEGRLPVMF
                PSGLIERPANWVALAGFMGTGKSRIGWELSRALALHFVDTDKLITRVVGKSIPEVFAQ
                EGEGYFRACEREVVQRVTRLDHAVISLGGGTFIHEENRRLLLSRGPVVVLWATPETVY
                QRTKHSDRPLLKVEDPLSRIRTLMNEREGVYRQGTIHVHSDGRPSEEIVEEVVERLWA
                WRDAQAAWAALPPDYDRATS"
gene            8965. .10017
                /gene="DR0777"
CDS             8965. .10017
                /gene="DR0777"
                /note="similar to PID:1653152 percent identity: 56.48;
                identified by sequence similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="3-dehydroquinate synthase"
                /protein_id="AAF10353.1"
                /db_xref="GI:6458486"
                /translation="MRRVEVGGAAPYAVEVGPGLLTRLAVPETPSALIHPADLPAAWV
                EQAQASLQPHLTVTVPARDECKTLEVFADVLSQLAQAGLPRSGAVVGLGGGAATDLAG
                FVAASYLRGVAFYTAPTTLLGMVDAAVGGKTGVNLPEGKNLVGAFWPPRAVWCDTATL
                TTLPDAVFREGAAEAFKHGLMADPSLLDRVVSPDFRPGGALLEDTLADAIAVKAGVVT
                RDLTEQGERAFLNFGHTLAHALEAVTHQAIPHGEAVAYGMHYAARLSHALGGADLTAH
                TRAFLRWQRPRRLPPLTFDALAPFIARDKKADSAGVRFVLLRDLAQPYLARVPDDVLR
                REFDGWQADLAELFPL"
BASE COUNT      1867 a   3523 c   3144 g   1587 t
ORIGIN


  Query Match             49.3%;  Score 22.2;  DB 1;  Length 10121;
  Best Local Similarity   77.1%;  Pred. No. 5.4e+02;
  Matches   27;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 aatccaagcttgcggccgatcaggccgaatatgcg 35
           || ||| ||| || ||||||||||| ||   |||
Db    2198 AAACCACGCTCGCCGCCGATCAGGCCAAAGCGGCG 2232


RESULT  13
AF012532
LOCUS       AF012532    18622 bp    DNA             BCT       12-AUG-1997
DEFINITION  Bacillus subtilis strain 168 trpC2 YefA (yefA) gene, partial cds,
            and YefB (yefB), YefC (yefC), YeeA (yeeA), YeeB (yeeB), YeeC
            (yeeC), YeeD (yeeD), YeeE (yeeE) and YeeF (yeeF) genes, complete
            cds.
ACCESSION   AF012532
VERSION     AF012532.1  GI:2318058
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 15361 to 18622)
  AUTHORS   Henriques,A.O., Beall,B.W., Roland,K. and Moran,C.P. Jr.
  TITLE     Characterization of cotJ, a sigma E-controlled operon affecting the
            polypeptide composition of the coat of Bacillus subtilis spores
  JOURNAL   J. Bacteriol. 177 (12), 3394-3406 (1995)
```

```
  MEDLINE   95286532
REFERENCE   2  (bases 1 to 18622)
  AUTHORS   Borriss,R. and Schroeter,R.
  TITLE     The 55-58 degree segment of the Bacillus subtilis chromosome, a
            region spanning from the purA gene cluster to the cotJ operon
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 18622)
  AUTHORS   Borriss,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1997) Borriss R., Humboldt University, Institute
            of Biology, Chausseestrasse 117, Berlin, Germany, D-10115
FEATURES             Location/Qualifiers
     source          1. .18622
                     /organism="Bacillus subtilis"
                     /strain="168 trpC2"
                     /db_xref="taxon:1423"
                     /map="55 to 58 degree"
     gene            <1. .81
                     /gene="yefA"
     CDS             <1. .81
                     /gene="yefA"
                     /note="similar to E.coli hypothetical RNA
                     methyltransferase: SwissProt Accession Number P55135;
                     similar to Chlamydia trachomatis Hom1: SwissProt Accession
                     Number P55137"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YefA"
                     /protein_id="AAB66480.1"
                     /db_xref="GI:2318067"
                     /translation="AKKRLRRSPPADNCGDEAEAGSVCVL"
     gene            complement(227. .1140)
                     /gene="yefB"
     CDS             complement(227. .1129)
                     /gene="yefB"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YefB"
                     /protein_id="AAB66472.1"
                     /db_xref="GI:2318059"
                     /translation="MYNEGVTSPSQLARKKNATDKWHGSSIRKILENPHYTGTLQQCR
                     EYRPSVTSKRRRSVNLENQIIIEKSHEPIIPLEDFLIVQEILITRKRRRPQAEVHLFT
                     NTAVCKDCRRSMHFKKNRKGYVCGSYNKHGLKACSDHYVSEIDLTDKVITNLNKIYLK
                     FSKENYFKELSEKALKYKEKIESKIIEINEKLNDKKRDKSNLVISLANGVISKEDYQL
                     AINITNEDISNFETTLHQLSKDLEFQKIEKEIIDFKKSLDKFMKEGTLTPEMLHLLVD
                     EIEVHANGTIEINYRFREPTVPSA"
     gene            1024. .1445
                     /gene="yefC"
     RBS             1024. .1030
                     /gene="yefC"
     CDS             1068. .1445
                     /gene="yefC"
                     /note="similar to Enterococcus faecalis plasmid PY117 orf4
                     encoded by GenBank Accession Number D78257"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YefC"
                     /protein_id="AAB66473.1"
                     /db_xref="GI:2318060"
                     /translation="MSVAFFLRASCEGLVTPSLYNPLAIASKPFPAIYSEKILIFTIF
                     SAFAEFERDMIVERTQEGKMLAKQNPDFREGRPKKFTKQQINHALTLLENHSYKQVED
                     MTGISVSTLVRAKKKKAAEAING"
     RBS             complement(1133. .1140)
                     /gene="yefB"
     gene            1505. .4159
                     /gene="yeeA"
     RBS             1505. .1512
                     /gene="yeeA"
     CDS             1520. .4159
                     /gene="yeeA"
                     /note="similar to Methanococcus janaschii plasmid pURB801
                     hypothetical protein MJECS02 encoded by GenBank Accession
                     Number L77119"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YeeA"
                     /protein_id="AAB66474.1"
                     /db_xref="GI:2318061"
                     /translation="MALIDLEDKIAEIVNREDHSDFLYELLGVYDVPRATITRLKKGN
                     QNLTKRVGEVHLKNKVWFKEAKKGKLFDALIDIEQQVEYLSAKPRYLLVTDYDGVLAK
                     DTKTLEALDVKFEELPQYFDFFLAWKGIEKVEFEKENPADIKAAERFARIYDVLRKEN
                     NIIETNRGLDLFLIRLLFCFFAEDTDIFKRNSFTNLIKTLTEEDGSNLNKLFADLFIV
                     LDKNERDDVPSYLKEFPYVNGQLFTEPHTELEFSAKSRKLIIECGELLNWAKINPDIF
                     GSMIQAVASEESRSYLGMHYTSVPNIMKVIKPLFLDKLNQSFLDAYDDYTKLENLLTR
                     IGKIKFFDPACGSGNFLIITYKELRRMEINIIKRLQELLGEYLYVPSVTLSQFYGIEI
                     EDFAHDVAKLSLWIAEHQMNEELKNEVHNAVRPTLPLHTAGDIRCANAIRVEWTEVCP
                     AQGSEEVYVFGNPPYLGSKKQNKEHKSDMLSIFGKVKNGKMLDYISAWFYFGAKYAST
                     TNAKVAFVSTNSVTQGEQVSILWNELFKFGIQINFANKSFKWANNAKNNAAVIVVIVG
                     FGPLDTKVNKYLFVDETKKLVSNISPYLTDGENILVSSRTKPISDLPKLHFGNMPNDG
                     GGLLFTITEYTDAINKYPELVPYFKKFIGSVEFINGGLRYCLWLNEAKYEKIKSNPLI
                     QERISISKNHREKSTDKGTNKLALTPWKFRDTHETTNYSIVVPSVSSENRFYIPMGLA
                     GADTILSNLIYVIYDAEIYLLGILMSRMHMTWVKAVAGRLKTDYRYSAGLCYNTFPIP
                     ELSTRRKNEIEEAILEILDLREEQGGTLAELYNPSTMPIELKVAHEKLDGIVERAYRQ
                     KQFESDEERLEVLLKLYQEMTER"
     RBS             4152. .4161
                     /gene="yeeB"
     gene            4152. .5967
                     /gene="yeeB"
     CDS             4171. .5967
                     /gene="yeeB"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YeeB"
                     /protein_id="AAB66475.1"
                     /db_xref="GI:2318062"
                     /translation="MENIIEINYNQTGKSKKTNEYGMREMQARAFEKRNSQYLLVKAP
                     PASGKSRALMFIGLDKLINQGLRKVIVAVPERSIGSSFKNTDLKSYGFFENWKVDPRN
                     NLTVGGDNSKVKSFVRFMESDDQVLICTHSTLRFAFEKIDDKAFDNCLLAIDEFHHVS
                     ADVNSRLGELLRSIIHNSSAHIVAMTGSYFRGDSVPILLPEDEELFDKVTYSYYEQLD
                     GYEYLKGFGIGYHIYQGQYTSAINEVLDTNKKTIVHIPNVNSGESTKDKYDEVGKILD
                     MIGEVEYQDDDTGIIYVKRHSDGKTLKIADLVDDQVGRENVVAYLRNIEALDDLDIII
                     ALGMAKEGFDWPFCEHTLTVGYRGSLTEIVQIIGRCTRDSYNKTYAQFTNLIAMPDAK
                     DEVVTYTVNTMLKAISASLLMEQVLTPDFKFKRRRNESEKSSTTGELFVKGLKEPSTE
                     NVKKIIENDINDLKAKTCKIHKSKTFSGEVDPKVLNKVLIPKVIQKVYPNLSDEEIEE
                     VRQHVVLDTVMKGAKSEVVGSKEFIRMADKFINIEDLDINLIDSINPFQKAFEVLSKE
                     LNSPVLKLIQETIDAKRINFDEDELVFIWPKV"
     RBS             6064. .6069
                     /gene="yeeC"
     gene            6064. .7228
                     /gene="yeeC"
     CDS             6080. .7228
                     /gene="yeeC"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YeeC"
                     /protein_id="AAB66476.1"
                     /db_xref="GI:2318063"
                     /translation="MASNRYHSINEIMESQLFYQITTPPKKTQKAQYDPEVEKFLEII
                     EFVKENGREPQKVPTDLTERSLASRLIGIRKDPDRMEYLKEYDEIGLLEVRSKELTIP
                     KISSIDDILKSGLSALLGDNLNNDITRSIFDTSSLQKVTTMPEYVAKRKKIKDFGKFE
                     ELFKKCHKEITEGKRKILTFKNEQDIQSNSFYILKGVLLYVEDVGERKKAKGKTNARL
                     RCIFENGTESDMLLRSLSAELYKHGRRVTDNEDTLLDNVREDDVSTGFIYVLKSLSTD
                     PQISSIKNLYKIGFTTGSVENRIRNAENQSTYLYAPVEIVTTYQVFNMNASKFETAIH
                     HALENNNLDVSILGANGKMLVPKEWFVVTLEDLQAVIDEIVMMVHLYD"
     gene            complement(7490. .8180)
                     /gene="yeeD"
     CDS             complement(7490. .7795)
                     /gene="yeeD"
                     /citation=[2]
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="YeeD"
                     /protein_id="AAB66477.1"
                     /db_xref="GI:2318064"
                     /translation="MTVRKELIKQINLTITVIKTINQKNPTPMVKNILKRYEEAKEFI
                     LQSSDGKFEEDLSKVRNKLDTLTRAYLESANDYMNPMLKEMHKTEKLLKEYDETTQS"
     RBS             complement(8175. .8180)
                     /gene="yeeD"
     terminator      8277. .8298
                     /note="rho independent terminator; deltaG= -10.2kcal/mol"
                     /citation=[2]
     gene            complement(8308. .9470)
                     /gene="yeeE"
     CDS             complement(8308. .9417)
                     /gene="yeeE"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="YeeE"
                     /protein_id="AAB66478.1"
                     /db_xref="GI:2318065"
                     /translation="MPAGVPYNAVNGVTLKEGLQKFAKVILPKPYGTSSSGRRTPAPH
                     VPPVTVKYGEHFARWSRKKVLKPNIIYKTKEGYTYTTDNYGRITSVKADLQLGEAKRN
                     QYAQTNAGKPQDRKPDDDGGHLIATQFKGSGQFDNIVPMNSQINRSGGKWYEMEQEWA
                     KALSKKPPKKVAVQIEPVYSGDSLRPSYFDVTYKIGSRKEISVSIKISLGVRRMETRK
                     MQDLYQLIGEKLNDIIPGEWTKIYLYAEVLDDSTMVLFHFRTPENNQIIYSQDIPSHY
                     NVSKDIFKTLLRELRELFEELRTEHRNNNDDVWTNLTLTLDRSGEFQLDYNYDDILAS
                     ELDGYERIAIWEYKNLGILPEDEDDKEFVISYLGL"
     RBS             complement(9464. .9470)
                     /gene="yeeE"
     gene            complement(9541. .10791)
                     /gene="yeeF"
     CDS             complement(9541. .10791)
                     /gene="yeeF"
                     /note="similar to Bacillus subtilis orf94 hypothetical
                     59.7 kDA protein in CWLA-CISA intergenic region: SwissProt
                     Accession Number P45942"

  Query Match             49.3%;  Score 22.2;  DB 1;  Length 18622;
  Best Local Similarity   69.8%;  Pred. No. 5e+02;
  Matches   30;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       3 tccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
           ||  || ||||||| | ||| || | |||| | || |||| |
Db   16961 TCAGAGTTTGCGGCTGCTCATGCTGCATATTGGACCTCATTTT 17003


RESULT  14
BSUB0004
LOCUS       BSUB0004    213190 bp    DNA             BCT       26-NOV-1997
DEFINITION  Bacillus subtilis complete genome (section 4 of 21): from 600701 to
            813890.
ACCESSION   Z99107 AL009126
VERSION     Z99107.1  GI:2632866
KEYWORDS    .
SOURCE      Bacillus subtilis.
  ORGANISM  Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE   1  (bases 1 to 213190)
  AUTHORS   Kunst,F., Ogasawara,N., Moszer,I., Albertini,A.M., Alloni,G.,
            Azevedo,V., Bertero,M.G., Bessieres,P., Bolotin,A., Borchert,S.,
            Borriss,R., Boursier,L., Brans,A., Braun,M., Brignell,S.C.,
            Bron,S., Brouillet,S., Bruschi,C.V., Caldwell,B., Capuano,V.,
            Carter,N.M., Choi,S.K., Codani,J.J., Connerton,I.F., Cummings,N.J.,
            Daniel,R.A., Denizot,F., Devine,K.M., Dusterhoft,A., Ehrlich,S.D.,
            Emmerson,P.T., Entian,K.D., Errington,J., Fabret,C., Ferrari,E.,
            Foulger,D., Fritz,C., Fujita,M., Fujita,Y., Fuma,S., Galizzi,A.,
            Galleron,N., Ghim,S.Y., Glaser,P., Goffeau,A., Golightly,E.J.,
            Grandi,G., Guiseppi,G., Guy,B.J., Haga,K., Haiech,J., Harwood,C.R.,
            Henaut,A., Hilbert,H., Holsappel,S., Hosono,S., Hullo,M.F.,
            Itaya,M., Jones,L., Joris,B., Karamata,D., Kasahara,Y.,
            Klaerr-Blanchard,M., Klein,C., Kobayashi,Y., Koetter,P.,
            Koningstein,G., Krogh,S., Kumano,M., Kurita,K., Lapidus,A.,
            Lardinois,S., Lauber,J., Lazarevic,V., Lee,S.M., Levine,A., Liu,H.,
            Masuda,S., Mauel,C., Medigue,C., Medina,N., Mellado,R.P.,
            Mizuno,M., Moestl,D., Nakai,S., Noback,M., Noone,D., O'Reilly,M.,
            Ogawa,K., Ogiwara,A., Oudega,B., Park,S.H., Parro,V., Pohl,T.M.,
            Portetelle,D., Porwollik,S., Prescott,A.M., Presecan,E., Pujic,P.,
            Purnelle,B., Rapoport,G., Rey,M., Reynolds,S., Rieger,M.,
            Rivolta,C., Rocha,E., Roche,B., Rose,M., Sadaie,Y., Sato,T.,
            Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J.,
            Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B.,
            Sorokin,A., Tacconi,E., Takagi,T., Takahashi,H., Takemaru,K.,
            Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpstra,P., Tognoni,A.,
            Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassarotti,A.,
            Viari,A., Wambutt,R., Wedler,E., Wedler,H., Weitzenegger,T.,
            Winters,P., Wipat,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K.,
            Yoshida,K., Yoshikawa,H.F., Zumstein,E., Yoshikawa,H. and
            Danchin,A.
    TITLE   The complete genome sequence of the gram-positive bacterium
            Bacillus subtilis
  JOURNAL   Nature 390 (6657), 249-256 (1997)
  MEDLINE   98044033
REFERENCE   2  (bases 1 to 213190)
  AUTHORS   Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.
    TITLE   Direct Submission
  JOURNAL   Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
            Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724
            Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
            adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
            68 89 48
FEATURES             Location/Qualifiers
     source          1. .213190
                     /organism="Bacillus subtilis"
                     /strain="168"
                     /db_xref="taxon:1423"
     gene            33. .740
                     /gene="ydfS"
     CDS             33. .740
                     /gene="ydfS"
                     /function="unknown"
                     /note="similar to hypothetical proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12373.1"
                     /db_xref="GI:2632867"
                     /db_xref="SPTREMBL:P96697"
                     /translation="MIELEVVIRTVASFGLLLIAERILGKQTISQMTIFDFIAAITLG
                     AIAAGLAYNTSIKPHNMAISFSIFVLTIFLISFLSIKNRKLRKFFAGDPTVLIQNGKI
                     LESNMRKMRYTLDYLNQQLREKEIFNIEEVLFAILETNGQLTVLRKPQFRHVTKQDLM
                     IAVNQEQRLPIELIMDGEIIENNLKQNRLTESWLLEELRKRDIKVKETVYAVLLGNGD
                     IYVDQYKDHISVPMDKE"
     gene            complement(755. .1186)
                     /gene="ydfT"
     CDS             complement(755. .1186)
                     /gene="ydfT"
                     /function="unknown"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12374.1"
                     /db_xref="GI:2632868"
                     /db_xref="SPTREMBL:P96698"
                     /translation="MDFEKIRKWLEITNEYKQSDFWTNVLKYKAPEHFFDSEASTFVY
                     DFYQDEEYNFIIVEMPGVYEEELTIRLLSKTQLLIKGTITPVFPAEMEVLRERYYGEI
                     ERIIQLPEAAETHLLQIQLLNGLLHISYPRQVETVAFNKGL"
     gene            complement(1199. .1441)
                     /gene="ydgA"
     CDS             complement(1199. .1441)
                     /gene="ydgA"
                     /function="unknown"
                     /note="similar to hypothetical proteins from B. subtilis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB12375.1"
                     /db_xref="GI:2632869"
```

```
                /db_xref="SPTREMBL:P96699"
                /translation="MPYQINIANIKINGVTQNGNIDVGPTVHNSHTANSKYFGANFSL
                GDLSPTSSLLNTGNIDSDVSDQDQIGNPSAPISNQI"
     gene       complement(1455. .1727)
                /gene="ydgB"
     CDS        complement(1455. .1727)
                /gene="ydgB"
                /function="unknown"
                /note="similar to hypothetical proteins from B. subtilis"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB12376.1"
                /db_xref="GI:2632870"
                /db_xref="SPTREMBL:P96700"
                /translation="MPSTVINLYYLKINSISGNGSITIGEAAYNSPTNNQKSQGTNSS
                FGDTSPTESVMENFLNDPDVNDQTSIGNSDTSNVNAPPIAPPPILD"
     gene       2026. .2613
                /gene="ydgC"
     CDS        2026. .2613
                /gene="ydgC"
                /function="unknown"
                /note="similar to hypothetical proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB12377.1"
                /db_xref="GI:2632871"
                /db_xref="SPTREMBL:P96701"
                /translation="MKSKGVESRKRLLKAAANEFSVRGFHDAKVSEIVKKAGFTQPSF
                YLYFQSKEAIFAELITDFHSRVRKLTESLLLENGLNTEDVSKRVLLAVETVFQFLDED
                KDLTKIGFFLNPEAKQMKKDLAMVLKENLEAEQRLGYFHSELDMETVAECLVGMIEHL
                TESFLLTGIKDPASLAAEVVNLLIYGMLPKGNDVR"
     gene       join(2610. .2954,3074. .3091)
                /gene="ydgD"
     CDS        2610. .2954
                /gene="ydgD"
                /function="unknown"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB12378.1"
                /db_xref="GI:2632872"
                /db_xref="SPTREMBL:P96702"
                /translation="MISIMMKVSLAVFMLAGGIIKVSRVPFQVEHWRHYQYPLWFLTV
                TGILEIAGALAMTAGIWNRYAAIGAGVLFVVLMAGAIHAHMFRARQSVIMAIQAMICL
                IVSIMIIMGSYT"
     terminator 3074. .3091
                /gene="ydgD"
     gene       3117. .3590
                /gene="ydgE"
     CDS        3117. .3590
                /gene="ydgE"
                /function="unknown"
                /note="similar to hypothetical proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB12379.1"
                /db_xref="GI:2632873"
                /db_xref="SPTREMBL:P96703"
                /translation="MAFTLEDMTEEEFEAFRGMSVQNFAKQNITSGTWTEKEAFEKSE
                QAYENMIPNGRDSSNHYFWNITNDQGERMGWLWLYADPLHPQKEAFIYSFGLYEAFRG
                KGLAQLALQTLDERARELGAERLALHVFAHNETAVYLYQKMGYAMTNIRMRKQLC"
     terminator complement(3705. .3722)
                /gene="expZ"
     gene       complement(3705. .5393)
                /gene="expZ"
     CDS        complement(3750. .5393)
                /gene="expZ"
                /codon_start=1
                /transl_table=11
                /product="ATP-binding transport protein"
                /protein_id="CAB12380.1"
                /db_xref="GI:2632874"
                /db_xref="SWISS-PROT:P39115"
                /translation="MKEIVTLTNVSYEVKDQTVFKHVNASVQQGDIIGIIGKNGAGKS
                TLLHLIHNDLAPAQGQILRKDIKLALVEQETAAYSFADQTPAEKKLLEKWHVPLRDFH
                QLSGGEKLKARLAKGLSEDADLLLLDEPTNHLDEKSLQFLIQQLKHYNGTVILVSHDR
                YFLDEAATKIWSLEDQTLIEFKGNYSGYMKFREKKRLTQQREYEKQQKMVERIEAQMN
                GLASWSEKAHAQSTKKEGFKEYHRVKAKRTDAQIKSKQKRLEKELEKAKAEPVTPEYT
                VRFSIDTTHKTGKRFLEVQNVTKAFGERTLFKNANFTIQHGEKVAIIGPNGSGKTTLL
                NIILGQETAEGSVWVSPSANIGYLTQEVFDLPLEQTPEELFENETFKARGHVQNLMRH
                LGFTAAQWTEPIKHMSMGERVKIKLMAYILEEKDVLILDEPTNHLDLPSREQLEETLS
                QYSGTLLAVSHDRYFLEKTTNSKLVISNNGIEKQLNDVPSERNEREELRLKLETERQE
                VLGKLSFMTPNDKGYKELDQAFNELTKRIKELDHQDKKD"
     terminator complement(5428. .5450)
                /gene="ydgF"
     gene       complement(5428. .7089)
                /gene="ydgF"
     terminator complement(5678. .5703)
                /gene="ydgF"
     CDS        complement(5713. .7089)
                /gene="ydgF"
                /function="unknown"
                /note="similar to amino acid ABC transporter (permease)"
                /codon_start=1
                /transl_table=11
                /protein_id="CAB12381.1"
                /db_xref="GI:2632875"
                /db_xref="SWISS-PROT:P96704"
                /translation="MTDDMTKDNINQQTLQRGLKNRHIQLIAIGGAIGTGLFLGSGKS
                IHFAGPSILFAYMITGIICFLIMRSLGELLLSNLNYHSFVDFVQDYLGDMAAFITGWT
                YWFCWISIAMADLTAVGLYTQYWLPGVPQWVPGLIALIILLIMNLATVKLFGELEFWF
                ALIKVIAILALIVIGLVMIFKGFSTSSGVSSFTNLWSHGGLFPNGMHGFILSFQMVVF
                AFVGIELVGLTAGETENPEKVIPKAINNIPVRVLLFYIGALLVIMSIYPWDIINPSES
                PFVQVFVAVGIVGAASIINFVVLTSAASACNSAVFSTSRMVYSLAKDHNAPESMAKLT
```

```
Query Match             49.3%;  Score 22.2;  DB 2;  Length 213190;
Best Local Similarity   69.8%;  Pred. No. 3.5e+02;
Matches   30;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      3 tccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
          ||  || ||||||| | ||| || | ||||  | || |||| |
Db 154535 TCAGAGTTTGCGGCTGCTCATGCTGCATATTGGACCTCATTTT 154577


RESULT  15
AE004219
LOCUS       AE004219    11936 bp    DNA             BCT       31-JUL-2000
DEFINITION  Vibrio cholerae chromosome I, section 127 of 251 of the complete
            chromosome.
ACCESSION   AE004219 AE003852
VERSION     AE004219.1  GI:9655899
KEYWORDS    .
SOURCE      Vibrio cholerae.
  ORGANISM  Vibrio cholerae
            Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE   1  (bases 1 to 11936)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     DNA Sequence of both chromosomes of the cholera pathogen Vibrio
            cholerae
  JOURNAL   Nature 406, 477-483 (2000)
REFERENCE   2  (bases 1 to 11936)
  AUTHORS   Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwinn,M.L.,
            Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
            Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
            Ermolaeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragoi,I.,
            Sellers,P., McDonald,L., Utterback,T., Fleishmann,R.D.,
            Nierman,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
            Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
```

```
FEATURES             Location/Qualifiers
     source          1. .11936
                     /organism="Vibrio cholerae"
                     /strain="N16961"
                     /serotype="O1"
                     /db_xref="taxon:666"
                     /chromosome="I"
                     /note="biotype: El Tor"
     gene            complement(98. .220)
                     /gene="VC1404"
     CDS             complement(98. .220)
                     /gene="VC1404"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF94561.1"
                     /db_xref="GI:9655900"
                     /translation="MENLEIQEQTIIKISTKNIDADLAQILFPILNKNKVNPCF"
     gene            460. .2247
                     /gene="VC1405"
     CDS             460. .2247
                     /gene="VC1405"
                     /note="similar to GB:X61367 PID:48195; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="methyl-accepting chemotaxis protein"
                     /protein_id="AAF94562.1"
                     /db_xref="GI:9655901"
                     /translation="MSIYLKALIPSRLRHFWSHILWHVTWNVRRTVMKWFNNLSITQK
                     MIALVGCLLALTLMVSTYSITKMKRVAVEIEAIAHDNLPLAKLMTDMTVHQLEGAITL
                     ERVFRAAGIESSQGKQQQNQYQAQLMELVNKFKQEFNQSRQLLESSIEHTISPSIRSY
                     LANLEQDLDSLGREHEEFERKLNELLNELRSGREVKLLVADAEQIEQLQISLDQRLID
                     ILRKIEQATEQSVLVTEQEEREALWGMIYLSAFAILFGLILGFAFSKQIVMAIARARK
                     LANEMAEGNFSKRAKVTTGDEIGQLITSMNTMAQSLSHIVGEVIDRANTIASTVTQLA
                     SSAESNKKSVQQQQANTEQVTSAMAQMAITITEVASSAEESSAATARAQENARYSCDV
                     LGKTETVSSQLVANAQQSQQMIVELEASTRQIESFVLVVEGISEQTNLLALNAAIEAA
                     RAGEQGRGFAVVADEVRALASRSQQATHEIKGLIQTLVERAQSATKMIDSSDRQIEES
                     FSSITAAKKQLDSINLALLELTSANTQVAAASEEQSVAADEISHNMTDIRDAGETIML
                     SAQETAQASEELAQQAQGLKLLMGRFVIS"
gene                 2326. .3642
                     /gene="VC1406"
CDS                  2326. .3642
                     /gene="VC1406"
                     /note="similar to GB:X61367 PID:48195; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="methyl-accepting chemotaxis protein"
                     /protein_id="AAF94563.1"
                     /db_xref="GI:9655902"
                     /translation="MIWKPFMPFWKKAAAQSEVEQHNHDSDIVSALKRNLAYIEFDPQ
                     GKILDANALFLSAMGYTKEEVIGQHHKIFCDPETVNSLEYAQFWKSLAQGQAQRKMFH
                     RLKKDGSSIWIEATYMPVRNRAGDVYKVIKVAHDVTKAKITADKQAAVFHSLDRSSAM
                     IQFNPDGTIKEANENFLKATGYRLEEIVGKHHRMFCDDRFYKENPNFWRELARGEFKS
                     GLFHRRTRHGQDLWLEATYNPIFNDAGVVTQVVKFASDVTDQVLKAHATKEASQMAQQ
                     TSAETVRVAESGREMIDAAATIASGITESIAGANALMTDLSSQSQRITQIVTTINKIA
                     EQTNLLALNAAIEAARAGEYGRGFAVVADEVRSLASNTSQATSEIDNIVKRNSQLTEQ
                     SGQTMEQIQAKVTEFNHMLQQTQSLIEQIQHAAENVQKTVSNIVDE"
gene                 3824. .5017
                     /gene="VC1407"
CDS                  3824. .5017
                     /gene="VC1407"
                     /note="similar to GP:3288677; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP-dependent RNA helicase RhlE"
                     /protein_id="AAF94564.1"
                     /db_xref="GI:9655903"
                     /translation="MSFSQLGLSDVLVQTVAQLGYQKPTHIQTQAIPVILQGRDLIAA
                     AQTGTGKTASFVLPILEKLRQGQTQRKKRVRALILVPTRELAMQVAEKVEQYGKDTGL
                     KSLAVFGGVDEQAQKQRLIDGVDVLVATPGRLMDLYGQRAVYFEEIEMVVLDEADRML
                     DMGFIESINKIIDCLPSEVQFLLFSATLSRKVRELAKTAVRDPHEISIAANQASKSNI
                     SQWLITVDKDTKSALLSHLINEQQWDQALIFIETKHGAAKLVSQLEKRGIHAEAFHSG
                     RSQAIRAQLLEDFKSGKIKYLVATGVAARGIDIDQLSRVVNYDLPFPADEYVHRIGRT
                     GRAEAVGEAISFVSKDNFKNLCMIESRLGHLIERRVVEGFEPKKPVPISILNYVPKHK
                     RAVQD"
gene                 complement(5076. .5678)
                     /gene="VC1408"
CDS                  complement(5076. .5678)
                     /gene="VC1408"
                     /note="similar to GB:M97208 SP:P32398 PID:143046
                     PID:2226233 GB:AL009126; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, TetR family"
                     /protein_id="AAF94565.1"
                     /db_xref="GI:9655904"
                     /translation="MRVKSEEKRQAILEIAKDSFTKQGFEQTSMSHIAKMLGGSKATL
                     YNYFSSKEEIFTAVMESSATEQIALSFLSLDHNRELKPALLEFGYNFLNSVLTQDAMS
                     IYRMAIHEADRSAIGRHFYENGPKRGWARLSRYITCQIECGSLKECDPWIAAMHFKAL
                     LSAQYFEPFSLGAIEKPTSSELKATTQRAVEAFLCLYAAD"
gene                 5785. .7239
                     /gene="VC1409"
CDS                  5785. .7239
                     /gene="VC1409"
                     /note="similar to GP:3378279; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="multidrug resistance protein, putative"
                     /protein_id="AAF94566.1"
                     /db_xref="GI:9655905"
                     /translation="MKNSVQTVGLLHKAAPYFTASLLSAFVLTGCSVPDHYPDLATMW
                     EANELSSTNTFSHQAEMDWPSANWWQRYQDAQLNHLIEEALQHSPSLEMAMARLKGAQ
                     GFARQAGAIRSFDLGLAASATESKVSERYQSATPPDGWNDYGTLTLNFQYDFDFWGKN
                     RAAVVAATSELAAAEAESVAARLMISTSIANAYAELARLYANQETVHAALQVRNKTVE
                     LLEKRYANGLETLGSVSQAKAVAASVEAELLGIQESIQLQKNALAALVGQGPDRAASI
                     EEPHITLTSRYGLPSEAGVGLLGHRADITAARWRAEAAAQQVGIAQAQFYPDVTLSAF
                     IGYQAFGLDHLFDSGNDAGAIGPAIYLPLFTGGRLEGQLTSAEARYQEAVAQYNGTLV
                     QALYEVADVVTSSQALQARINKTEQAVQQAEQALHIATNRYQGGLATYLDVLVAEESL
                     LNNQRALVNLQSRAFSLDLALIHALGGGFETTES"
gene                 7220. .8440
                     /gene="VC1410"
CDS                  7220. .8440
                     /gene="VC1410"
                     /note="similar to GP:3378278; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="multidrug resistance protein VceA"
                     /protein_id="AAF94567.1"
                     /db_xref="GI:9655906"
                     /translation="MKQQNLNRLSSMNSNNSNTEFDGVRAAQSRKKGFLGLAAAIVVA
                     GGSYALYWHFIGSRYISTDNAYAAAEIAEVTPAVGGIIAQVNVVDTEYVKQGDVLVQL
                     DDTDARLALLQAEADLALAKRRVRSYLANDEGLSAMVEAQEANEQRVKAQLKAAQADF
                     ERAKIDLSRREDLVRSGSVSGEELTNAKTGFAQAQANLNAAKAAMAQAQATKLSTIGS
                     QKANAALTDNTTVDSNPEVLLAKARYEQAKIDLERTVIRAPISGIVAKRQVQVGRRVQ
                     VGMPLMTVVPTDHIYVDANFKEVELRDVKVGQPVTLTADLYGDDVTYHGVVAGFSGGT
                     GSAFSMIPAQNATGNWIKVVQRLPIRIELDPKDLQAYPLQVGLSMVATIDTAGTTDPQ
                     TLVQYRAAKVSEQG"
gene                 8450. .9985
                     /gene="VC1411"
CDS                  8450. .9985
                     /gene="VC1411"
                     /note="similar to GP:2815578; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="multidrug resistance protein VceB"
                     /protein_id="AAF94568.1"
                     /db_xref="GI:9655907"
                     /translation="MSHNADNEMQPLSGWALFFGALCLAMANFLAILDTTIANVSVSN
```

```
                     IAGSLGTSTSQGTYVITSYAVAEAISVPLTGWLASRFGSIRVFVTCFLLFGVFSLLCG
                     LANSMSTLVMFRVLLGFVGGPLMPLSQTLMIRIFPKNKSHAAIGIWSMTTLVAPIMGP
                     ILGGVLCDQLSWPYIFFIKMPFAIAAALLCWKLLKKFETKTTHSKIDKVGLALLVVWV
                     AALQLMLDEGKDHDWFESSRIVFLAVIAVIGFIAFLIWELTERNPVVDLKVFRHRGYS
                     ISMVTLSLAFGAFFSISVVTPLWLQIYMGYTATISGHATASMGILAVFLAPIVANLSS
                     KFDPRPFVFAGVMWLGLWTFMRGFNTVDMTFSQISWPLFFQGIGMPLFFVPLTAIALG
                     SVKPHEMESAAGLMNFIRTLSGAFATSMINTSWEHETRYVHAELAGLTDKAGVATQAM
                     QNSGMSAEQTRSAMDWILQNQSVMVATNQLFIVIALIFVFAACMIWFAPKPKQAVDTS
                     AVH"
     gene            9939. .10055
                     /gene="VC1412"
     CDS             9939. .10055
                     /gene="VC1412"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAF94569.1"
                     /db_xref="GI:9655908"
                     /translation="MRQNPSKRWIPLLFIKRQRRNTHGITPLFSRAYECGED"
     gene            complement(10060. .11796)
                     /gene="VC1413"
     CDS             complement(10060. .11796)

  Query Match          48.9%;  Score 22;  DB 1;  Length 11936;
  Best Local Similarity  73.7%;  Pred. No. 6.2e+02;
  Matches   28;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       3 tccaagcttgcggccgatcaggccgaatatgcggccgc 40
           ||| ||||||| |||| | |||| ||    || ||| ||
Db    6289 TCCGAGCTTGCTGCCGCTGAGGCAGAGAGTGTGGCTGC 6326


Search completed: March  4, 2001, 12:03:42
Job time: 5597 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 10:32:15 ; Search time 106.76 Seconds
(without alignments)
67.930 Million cell updates/sec

Title: US-09-101-423B-7
Perfect score: 45
Sequence: 1 aatccaagcttgcggccgat.....cgaatatgcggccgcattat 45

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	21.2	47.1	3481	3	US-09-136-251-1	Sequence 1, Appli
c	2	20.4	45.3	41	1	US-08-200-900A-26	Sequence 26, Appl
c	3	20.4	45.3	41	4	PCT-US94-00616-26	Sequence 26, Appl
	4	20.2	44.9	2885	3	US-08-948-705-2	Sequence 2, Appli
c	5	19.8	44.0	3415	1	US-08-054-077C-1	Sequence 1, Appli
	6	19	42.2	638	1	US-08-469-667-1	Sequence 1, Appli
	7	19	42.2	638	4	PCT-US95-07289-1	Sequence 1, Appli
	8	18.8	41.8	38506	3	US-09-320-878-19	Sequence 19, Appl
	9	18.6	41.3	37	2	US-08-839-581A-21	Sequence 21, Appl
	10	18.6	41.3	62	5	5332671-12	Patent No. 5332671
c	11	18.6	41.3	1708	4	PCT-US95-02481-3	Sequence 3, Appli
c	12	18.6	41.3	8051	2	US-08-576-626A-2	Sequence 2, Appli
	13	18.4	40.9	88	1	US-08-706-037-16	Sequence 16, Appl
	14	18.4	40.9	88	2	US-09-005-397-16	Sequence 16, Appl
	15	18.4	40.9	68750	3	US-09-335-409-1	Sequence 1, Appli
	16	18.2	40.4	1374	1	US-08-278-630A-9	Sequence 9, Appli
	17	18.2	40.4	1424	1	US-08-403-634-3	Sequence 3, Appli
	18	18.2	40.4	1424	3	US-08-913-441B-3	Sequence 3, Appli
	19	18.2	40.4	1584	3	US-08-907-229-1	Sequence 1, Appli
	20	18.2	40.4	1965	1	US-08-258-420-9	Sequence 9, Appli
	21	18.2	40.4	2001	3	US-08-850-961-13	Sequence 13, Appl
	22	18.2	40.4	2943	1	US-08-042-747A-7	Sequence 7, Appli
	23	18.2	40.4	6028	3	US-09-011-745-5	Sequence 5, Appli
c	24	18	40.0	495	1	US-08-406-248-1	Sequence 1, Appli
c	25	18	40.0	1194	4	PCT-US94-09700-10	Sequence 10, Appl
	26	18	40.0	1721	1	US-08-241-766-3	Sequence 3, Appli
c	27	18	40.0	2100	1	US-08-154-915-5	Sequence 5, Appli
c	28	18	40.0	2106	1	US-07-970-462A-1	Sequence 1, Appli
c	29	18	40.0	2106	4	PCT-US92-10904-1	Sequence 1, Appli
c	30	18	40.0	2106	4	PCT-US94-09700-1	Sequence 1, Appli
c	31	18	40.0	2121	1	US-08-574-043A-1	Sequence 1, Appli
c	32	18	40.0	2121	2	US-08-795-015-1	Sequence 1, Appli
c	33	18	40.0	2121	4	PCT-US94-12936-1	Sequence 1, Appli
c	34	18	40.0	2121	4	PCT-US96-11886A-1	Sequence 1, Appli
c	35	18	40.0	2245	2	US-08-617-801A-3	Sequence 3, Appli
	36	18	40.0	3051	1	US-08-241-766-10	Sequence 10, Appl
c	37	18	40.0	6914	1	US-08-920-812-22	Sequence 22, Appl
c	38	18	40.0	6914	1	US-08-920-827-22	Sequence 22, Appl
c	39	18	40.0	6914	1	US-08-921-177-22	Sequence 22, Appl
c	40	18	40.0	6914	1	US-08-362-577C-22	Sequence 22, Appl
c	41	18	40.0	6914	2	US-08-920-828-22	Sequence 22, Appl
	42	17.8	39.6	981	1	US-08-259-148A-8	Sequence 8, Appli
	43	17.8	39.6	981	1	US-08-484-054-8	Sequence 8, Appli
	44	17.8	39.6	981	2	US-07-876-941A-8	Sequence 8, Appli
	45	17.8	39.6	984	1	US-08-240-049B-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-136-251-1/c
; Sequence 1, Application US/09136251A
; Patent No. 6127156
; GENERAL INFORMATION:
; APPLICANT: HOSHINO, Tatsuo
; APPLICANT: MIYAZAKI, Taro
; APPLICANT: OJIMA, Setsuko
; APPLICANT: SHINJOH, Masako
; APPLICANT: TOMIYAMA, NO. 6127156ibumi
; TITLE OF INVENTION: D-SORBITOL DEHYDROGENASE GENE
; FILE REFERENCE: D-Sorbital Dehydrogenase Gene
; CURRENT APPLICATION NUMBER: US/09/136,251A
; CURRENT FILING DATE: 1998-08-19
; EARLIER APPLICATION NUMBER: EP 97114432.4
; EARLIER FILING DATE: 1997-08-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3481
; TYPE: DNA
; ORGANISM: Gluconobacter suboxydans
US-09-136-251-1

Query Match 47.1%; Score 21.2; DB 3; Length 3481;
Best Local Similarity 76.5%; Pred. No. 11;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy       4 ccaagcttgcggccgatcaggccgaatatgcggc 37
           |||||| ||||||| || ||   ||||| || ||
Db     388 CCAAGCGTGCGGCCCATGAGCATGAATACGCCGC 355
```

RESULT 2
US-08-200-900A-26/c
; Sequence 26, Application US/08200900A
; Patent No. 5665566
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. - Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/200,900A
;      FILING DATE:  23-FEB-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meinert, Maureen C.
;      REGISTRATION NUMBER:  31,544
;      REFERENCE/DOCKET NUMBER:  GI 5201-FWC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 876-1170 X8574
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  41 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-200-900A-26


  Query Match             45.3%;  Score 20.4;  DB 1;  Length 41;
  Best Local Similarity   71.1%;  Pred. No. 13;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       8 gcttgcggccgatcaggccgaatatgcggccgcattat 45
           |||||      | ||||| |||  |||||||||||| |
Db      41 GCTTGGACTTAACCAGGCTGAACTTGCGGCCGCATTCT 4


RESULT   3
PCT-US94-00616-26/c
; Sequence 26, Application PC/TUS9400616
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  CLONING OF ENTEROKINASE AND METHOD OF USE
;    NUMBER OF SEQUENCES:  33
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/00616
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  41 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US94-00616-26


  Query Match             45.3%;  Score 20.4;  DB 4;  Length 41;
  Best Local Similarity   71.1%;  Pred. No. 13;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       8 gcttgcggccgatcaggccgaatatgcggccgcattat 45
           |||||      | ||||| |||  |||||||||||| |
Db      41 GCTTGGACTTAACCAGGCTGAACTTGCGGCCGCATTCT 4


RESULT   4
US-08-948-705-2
; Sequence 2, Application US/08948705A
; Patent No. 6043084
; GENERAL INFORMATION:
;  APPLICANT: Scanlan, Matthew J.
;  APPLICANT: Chen, Yao-Tseng
;  APPLICANT: Stockert, Elisabeth
;  APPLICANT: Old, Lloyd J.
;  TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
;  TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND METHODS FOR DIAGNOSING AND
;  TITLE OF INVENTION: TREATING COLON CANCER
;  FILE REFERENCE: LUD-5506-JEL/NDH
;  CURRENT APPLICATION NUMBER: US/08/948,705A
;  CURRENT FILING DATE: 1997-10-10
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2885
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
US-08-948-705-2


  Query Match             44.9%;  Score 20.2;  DB 3;  Length 2885;
  Best Local Similarity   75.8%;  Pred. No. 25;
  Matches   25;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       8 gcttgcggccgatcaggccgaatatgcggccgc 40
           | | | |||||| |||||| ||  |||||| ||
Db    2715 ggtgggggccgaacaggcccaagctgcggcagc 2747


RESULT   5
US-08-054-077C-1/c
; Sequence 1, Application US/08054077C
; Patent No. 5527679
;  GENERAL INFORMATION:
;    APPLICANT:  HEMLER, MARTIN E.
;    APPLICANT:  RAMASWAMY, HEMAVATHI
;    TITLE OF INVENTION:  HUMAN INTEGRIN  5 SUBUNIT PROTEIN
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
;      ADDRESSEE:  CUSHMAN
;      STREET:  130 WATER STREET
;      CITY:  BOSTON
;      STATE:  MASSACHUSETTS
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/054,077C
;      FILING DATE:  27-APR-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/694314
;      FILING DATE:  01-MAY-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RESNICK, DAVID S.
;      REGISTRATION NUMBER:  34235
;      REFERENCE/DOCKET NUMBER:  40937
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 523-3400
;      TELEFAX:  (617) 523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3415 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  mat peptide
```

```
;       LOCATION:  406..2733
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  337..2733
US-08-054-077C-1


  Query Match             44.0%;  Score 19.8;  DB 1;  Length 3415;
  Best Local Similarity   69.2%;  Pred. No. 36;
  Matches   27;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      4 ccaagcttgcggccgatcaggccgaatatgcggccgcat 42
          ||| ||| || || |||| ||||   | |||  | ||||
Db   2153 CCACGCTCGCTGCAGATCTGGCCATCTCTGCCCCGGCAT 2115


RESULT   6
US-08-469-667-1
; Sequence 1, Application US/08469667
; Patent No. 5733748
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Guo-Liang
;    APPLICANT:  Rosen, Craig
;    TITLE OF INVENTION:  Colon Specific Genes and Proteins
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi,
;      ADDRESSEE:  Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07068-1739
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/469,667
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferraro, Gregory D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-435
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  638 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..501
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..501
US-08-469-667-1


  Query Match             42.2%;  Score 19;  DB 1;  Length 638;
  Best Local Similarity   75.9%;  Pred. No. 60;
  Matches   22;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      7 agcttgcggccgatcaggccgaatatgcg 35
          |||| || |||:| |||||||  |||| |
Db      8 AGCTGGCTGCCSACCAGGCCGTGTATGTG 36


RESULT   7
PCT-US95-07289-1
; Sequence 1, Application PC/TUS9507289
;  GENERAL INFORMATION:
;    APPLICANT:  Yu, Guo-Liang
;    APPLICANT:  Rosen, Craig
;    TITLE OF INVENTION:  Colon Specific Genes and Proteins
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carella, Byrne, Bain, Gilfillan, Cecchi,
;      ADDRESSEE:  Stewart & Olstein
;      STREET:  6 Becker Farm Road
;      CITY:  Roseland
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07068-1739
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07289
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferraro, Gregory D.
;      REGISTRATION NUMBER:  36,134
;      REFERENCE/DOCKET NUMBER:  325800-265
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201-994-1700
;      TELEFAX:  201-994-1744
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  638 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1..501
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..501
PCT-US95-07289-1


  Query Match             42.2%;  Score 19;  DB 4;  Length 638;
  Best Local Similarity   75.9%;  Pred. No. 60;
  Matches   22;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      7 agcttgcggccgatcaggccgaatatgcg 35
          |||| || |||:| |||||||  |||| |
Db      8 AGCTGGCTGCCSACCAGGCCGTGTATGTG 36


RESULT   8
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
```

; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match 41.8%; Score 18.8; DB 3; Length 38506;
Best Local Similarity 68.4%; Pred. No. 1.1e+02;
Matches 26; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 8 gcttgcggccgatcaggccgaatatgcggccgcattat 45
| | ||||| | |||||||| | ||||||| | |
Db 1851 gttcgcggcggccatggccgaatgcgaggccgcactct 1888

RESULT 9
US-08-839-581A-21
; Sequence 21, Application US/08839581A
; Patent No. 5958705
; GENERAL INFORMATION:
; APPLICANT: Staunton, Donald E.
; APPLICANT: Lipsky, Brian P.
; TITLE OF INVENTION: Cytoplasmic Modulators of Integrin
; TITLE OF INVENTION: Binding/Signaling
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,581A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,337
; REFERENCE/DOCKET NUMBER: 27866/33886
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 37 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "primer"
US-08-839-581A-21

Query Match 41.3%; Score 18.6; DB 2; Length 37;
Best Local Similarity 84.0%; Pred. No. 60;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 atccaagcttgcggccgatcaggcc 26
|| |||||||||||| ||||| |
Db 1 ATATAAGCTTGCGGCCGCTCAGGGC 25

RESULT 10
5332671-12
;Patent No. 5332671
; APPLICANT: FERRARA, NAPOLEONE;LEUNG, DAVID W.H.
; TITLE OF INVENTION: PRODUCTION OF VASCULAR ENDOTHELIAL CELL
;GROWTH FACTOR AND DNA ENCODING SAME
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/389,722
; FILING DATE: 04-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 369,424
; FILING DATE: 21-JUN-1989
; APPLICATION NUMBER: 351,117
; FILING DATE: 12-MAY-1989
;SEQ ID NO:12:
; LENGTH: 62
5332671-12

Query Match 41.3%; Score 18.6; DB 5; Length 62;
Best Local Similarity 72.7%; Pred. No. 64;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 aatccaagcttgcggccgatcaggccgaatatg 33
||| |||||||||||||| | || | || |
Db 12 aattcaagcttgcggccgctaagagctcttaag 44

RESULT 11
PCT-US95-02481-3/c
; Sequence 3, Application PC/TUS9502481
; GENERAL INFORMATION:
; APPLICANT: Fioretti, William C.
; APPLICANT: Kousoulas, Konstantin
; APPLICANT: Satterlee, Daniel G.
; TITLE OF INVENTION: Methods of Determining the Gender of Poultry
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew
; STREET: 191 Peachtree Street, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1769
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50
; COMPUTER: MacIntosh
; OPERATING SYSTEM: 7.0
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02481
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/202,909
; FILING DATE: February 28, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Roxanne Edwards Cenatempo
; REGISTRATION NUMBER: 38,767
; REFERENCE/DOCKET NUMBER: 01051-0101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799

```
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1708 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US95-02481-3


  Query Match             41.3%;  Score 18.6;  DB 4;  Length 1708;
  Best Local Similarity   65.9%;  Pred. No. 95;
  Matches   27;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      4 ccaagcttgcggccgatcaggccgaatatgcggccgcatta 44
          ||  || ||| || ||||| ||  || ||   || | |||
Db    260 CCCTGCCTGCAGCAGATCAGTCCACATCTGTCTCCTCGTTA 220


RESULT  12
US-08-576-626A-2/c
; Sequence 2, Application US/08576626A
; Patent No. 5998194
;  GENERAL INFORMATION:
;    APPLICANT:  Summers, R.G.
;    APPLICANT:  Katz, L.
;    APPLICANT:  Donadio, S.
;    APPLICANT:  Staver, M.J.
;    TITLE OF INVENTION:  POLYKETIDE-ASSOCIATED SUGAR
;    TITLE OF INVENTION:  BIOSYNTHESIS GENES
;    NUMBER OF SEQUENCES:  60
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/576,626A
;      FILING DATE:  21-DEC-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  5857.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847) 938-3137
;      TELEFAX:  (847) 938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8051 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-576-626A-2


  Query Match             41.3%;  Score 18.6;  DB 2;  Length 8051;
  Best Local Similarity   72.7%;  Pred. No. 1.1e+02;
  Matches   24;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy     11 tgcggccgatcaggccgaatatgcggccgcatt 43
          |||||||| |||| |||  | | || || | |
Db   6190 TGCGGCCGAACAGGTCGAGCAGGTGGGCGAACT 6158


RESULT  13
US-08-706-037-16
; Sequence 16, Application US/08706037
; Patent No. 5770419
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Feng
;    APPLICANT:  Berka, Randy M.
;    APPLICANT:  Wahleithner, Jill A.
;    TITLE OF INVENTION:  BLUE COPPER OXIDASE MUTANTS WITH
;    TITLE OF INVENTION:  ENHANCED ACTIVITY
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5770419o No. 5770419disk of No. 5770419th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/706,037
;      FILING DATE:  30-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/002,800
;      FILING DATE:  1-SEP-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Harrington, James J.
;      REGISTRATION NUMBER:  38,711
;      REFERENCE/DOCKET NUMBER:  4526.200-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  88 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-706-037-16


  Query Match             40.9%;  Score 18.4;  DB 1;  Length 88;
  Best Local Similarity   63.6%;  Pred. No. 79;
  Matches   28;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      1 aatccaagcttgcggccgatcaggccgaatatgcggccgcatta 44
            ||| ||||||||  | || | |  || ||| || |  | | | |
Db     20 AATTCAAGCTTGTCGACGTTTAAACCAAATTTGAGATCTCTTAA 63


RESULT  14
US-09-005-397-16
; Sequence 16, Application US/09005397
; Patent No. 5972670
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Feng
;    APPLICANT:  Berka, Randy M.
;    APPLICANT:  Wahleithner, Jill A.
;    TITLE OF INVENTION:  BLUE COPPER OXIDASE MUTANTS WITH
;    TITLE OF INVENTION:  ENHANCED ACTIVITY
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5972670o No. 5972670disk of No. 5972670th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
```

; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,397
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/706,037
; FILING DATE: 30-AUG-1996
; APPLICATION NUMBER: US 60/002,800
; FILING DATE: 1-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4526.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 88 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
-US-09-005-397-16

Query Match 40.9%; Score 18.4; DB 2; Length 88;
Best Local Similarity 63.6%; Pred. No. 79;
Matches 28; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 aatccaagcttgcggccgatcaggccgaatatgcggccgcatta 44
||| |||||||| | || | | || ||| || | | | | |
Db 20 AATTCAAGCTTGTCGACGTTTAAACCAAATTTGAGATCTCTTAA 63

RESULT 15
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 40.9%; Score 18.4; DB 3; Length 68750;
Best Local Similarity 78.6%; Pred. No. 1.7e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 13 cggccgatcaggccgaatatgcggccgc 40
|||||| ||| || | |||||||| ||

Db 36360 cggccgggcagtccaactatgcggcggc 36387

Search completed: March 4, 2001, 12:05:17
Job time: 5582 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 10:32:45 ; Search time 96.53 Seconds
(without alignments)
175.125 Million cell updates/sec

Title: US-09-101-423B-7
Perfect score: 45
Sequence: 1 aatccaagcttgcggccgat.....cgaatatgcggccgcattat 45

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36:*
1: /cgn2_2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	45	100.0	45	18	T72784	DNA tag. Syntheti
	2	21.4	47.6	42	18	T61221	Primer HUCKN.NOT f
c	3	21.2	47.1	500	20	X91235	T. gondii immunoge
c	4	21.2	47.1	3481	20	X21501	DNA sequence of so
	5	20.8	46.2	1503	19	V15518	Hamster oral papil
	6	20.8	46.2	7647	19	V15519	Hamster oral papil
	7	20.8	46.2	8907	21	Z36926	Nucleotide sequenc
	8	20.8	46.2	9181	20	X84323	Stealth virus nucl
c	9	20.4	45.3	41	15	Q70114	Lambda INNER prime
	10	20.4	45.3	939	21	Z29251	Rifin 3193 gene.
c	11	20.4	45.3	4260	9	N81768	Sequence encoding
c	12	20.4	45.3	4380	9	N80222	Sequence of Mycoba
c	13	20.4	45.3	4380	19	V05708	Mycobacterium tube
c	14	20.2	44.9	1818	21	Z46842	Semaphorin K1 gene
	15	20.2	44.9	2885	20	X40055	Colon cancer assoc
	16	20.2	44.9	5398	12	Q13284	P.denitrificans co
c	17	20	44.4	2178	20	Z12261	Neisseria gonorrho
c	18	19.8	44.0	3415	17	T36481	Human integrin bet
c	19	19.8	44.0	7673	19	V58229	Omega-cyclohexane
c	20	19.6	43.6	3183	20	X26611	Trypanosoma cruzi
c	21	19.4	43.1	424	20	X84148	DNA encoding human
	22	19.4	43.1	524	21	Z53380	Neisseria gonorrhe
	23	19.4	43.1	546	20	Z12160	Neisseria gonorrho
	24	19.4	43.1	546	21	Z53377	Neisseria gonorrhe
c	25	19.4	43.1	753	21	Z53374	Neisseria gonorrhe
	26	19.4	43.1	2352	21	Z54296	Neisseria meningit
	27	19.2	42.7	619	21	Z53637	Neisseria gonorrhe
	28	19.2	42.7	834	21	Z53639	Neisseria meningit
	29	19.2	42.7	1095	21	Z54498	Neisseria meningit
	30	19.2	42.7	1842	21	Z56349	Escherichia coli f
	31	19	42.2	417	20	V90226	EST clone DF860.
	32	19	42.2	638	18	T45880	Human colon specif
	33	19	42.2	638	19	V16668	Polynucleotide seq
	34	19	42.2	1173	20	X91764	Porphorymonas ging
	35	19	42.2	1218	20	X91630	Porphorymonas ging
	36	19	42.2	4394	13	Q21604	Alpha galactosidas
c	37	19	42.2	5474	16	Q90251	Tuberous sclerosis
c	38	18.8	41.8	2302	21	A15621	Human heat shock p
c	39	18.8	41.8	2302	21	A15622	Human heat shock p
c	40	18.8	41.8	2465	18	T58086	Human heat shock p
c	41	18.8	41.8	2465	21	A15620	Human heat shock p
c	42	18.8	41.8	4248	21	A27364	Sinorhizobium meli
c	43	18.8	41.8	7059	15	Q55004	Humanised anti-CD1
c	44	18.8	41.8	7812	20	X12962	Enterococcus faeca
	45	18.8	41.8	13842	21	Z87297	S. venezuelae macr

ALIGNMENTS

RESULT 1
T72784
ID T72784 standard; DNA; 45 BP.
XX
AC T72784;
XX
DT 22-SEP-1997 (first entry)
XX
DE DNA tag.
XX
KW Metastasis-inducing DNA; Met-DNA; cancer; diagnosis; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 1..2
FT /*tag= a
FT /note= "5' single stranded overhang"
FT misc_feature 45
FT /*tag= b
FT /note= "5' overhang on complementary strand of 4
FT bases with sequence 5'-AGCT-3'"
XX
PN WO9725443-A1.
XX
PD 17-JUL-1997.
XX
PF 10-JAN-1997; 97WO-GB00074.
XX
PR 10-JAN-1996; 96GB-0000470.
XX
PA (UYLI-) UNIV LIVERPOOL.
XX
PI Barraclough BR, Rudland PS;
XX

```
DR   WPI; 1997-372878/34.
XX
PT   New isolated metastasis-inducing DNA - used to develop products to
PT   identify and treat patients at risk from metastatic tumours
XX
PS   Claim 6; Page 24; 38pp; English.
XX
CC   A DNA tag (T72784) having HindIII, SfiI, NotI and a defective
CC   HindIII site is used to tag human DNA from malignant, metastatic
CC   cancer cells in a novel method of screening and recovering
CC   metastasis-inducing DNA (Met-DNA).  This method involves
CC   transferring the human DNA into a cell line (pref. rat Rama 37)
CC   that produces only benign, non-metastasing tumours when injected
CC   into a syngeneic animal, injecting the transformed cells into a
CC   syngeneic animal, selecting those animals having metastasing
CC   tumours, and recovering the Met-DNA (see also T72785-90) from them.
CC   The human DNA is tagged to assist in its removal or insertion from
CC   or into a host or vector.  The tagging procedure overcomes the
CC   problem of identifying the inserted human DNA sequences in the rat
CC   genome of the transfected cells.
XX
SQ   Sequence 45 BP; 11 A; 13 C; 12 G; 9 T; 0 other;


  Query Match             100.0%;  Score 45;  DB 18;  Length 45;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-09;
  Matches   45;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aatccaagcttgcggccgatcaggccgaatatgcggccgcattat 45
          |||||||||||||||||||||||||||||||||||||||||||||
Db      1 aatccaagcttgcggccgatcaggccgaatatgcggccgcattat 45


RESULT   2
T61221
ID   T61221 standard; DNA; 42 BP.
XX
AC   T61221;
XX
DT   17-OCT-1997  (first entry)
XX
DE   Primer HUCKN.NOT for light chain constant region cDNA.
XX
KW   Primer; polymerase chain reaction; PCR; amplification; light chain;
KW   immunoglobulin; Ig; constant; region; production;
KW   recombinant; antibody; B cell; diagnosis; therapy; ss.
XX
OS   Synthetic.
XX
PN   DE19526546-A1.
XX
PD   23-JAN-1997.
XX
PF   20-JUL-1995;   95DE-1026546.
XX
PR   20-JUL-1995;   95DE-1026546.
XX
PA   (OPEL/) OPELZ G.
XX
PI   Terness P,  Welschof M;
XX
DR   WPI; 1997-088250/09.
XX
PT   Prodn. of recombinant antibodies - by amplification, cloning and
PT   expression of cDNA generated from B-cell mRNA
XX
PS   Claim 5; Page 11; 11pp; German.
XX
CC   The present sequence is a primer for the PCR amplification of the
CC   cDNA encoding an immunoglobulin (Ig), light chain, constant region.
CC   The primer can be used in a novel method for the production of
CC   recombinant antibodies, comprising the selection of B cells from a
CC   lymphocyte fraction, isolation of mRNA from individual B cells,
CC   reverse transcription of the mRNA into cDNA, amplification of the
CC   cDNA by PCR and cloning and expression of the cDNA. The recombinant
CC   antibodies can be used for diagnosis and/or therapy, while the
CC   method avoids the need for intermediate separation of light and
CC   heavy chains and gene library screening.
XX
SQ   Sequence 42 BP; 11 A; 11 C; 14 G; 6 T; 0 other;


  Query Match             47.6%;  Score 21.4;  DB 18;  Length 42;
  Best Local Similarity   71.8%;  Pred. No. 9.2;
  Matches   28;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      5 caagcttgcggccgatcaggccgaatatgcggccgcatt 43
          |||||||||||||||   || | ||   ||| ||| || |
Db      4 caagcttgcggccgcgaagacagatggtgcagccacagt 42


RESULT   3
X91235/c
ID   X91235 standard; DNA; 500 BP.
XX
AC   X91235;
XX
DT   24-SEP-1999  (first entry)
XX
DE   T. gondii immunogenic protein encoding DNA.
XX
KW   Immunogenic protein; Toxoplasma gondii protein; oocyst shedding; cat;
KW   T. gondii infection; enteric apicomplexa oocyst; Cryptosporidium oocyst;
KW   Toxoplasma oocyst; ss.
XX
OS   Toxoplasma gondii.
XX
PN   WO9932633-A1.
XX
PD   01-JUL-1999.
XX
PF   18-DEC-1998;   98WO-US27137.
XX
PR   19-DEC-1997;   97US-0994825.
XX
PA   (HESK-) HESKA CORP.
XX
PI   Lutz SB,  Milhausen MJ,  Ng RK;
XX
DR   WPI; 1999-418930/35.
XX
PT   New isolated Toxoplasma gondii nucleic acids used, e.g. to treat
PT   infection caused by this microorganism
XX
PS   Claim 4; Page 213; 381pp; English.
XX
CC   The invention provides isolated Toxoplasma gondii nucleic acids that
CC   encode immunogenic polypeptides. The T. gondii nucleic acid molecules,
CC   immunogenic proteins and antibodies to the proteins can be used to
CC   inhibit T. gondii oocyst shedding in a cat due to infection with
CC   T. gondii. They can be used for preventing T. gondii infection and for
CC   preventing the spread of T. gondii infection. They can also be used for
CC   detecting T. gondii infection. The detection method can be used to detect
CC   parasite cysts or oocysts in feces, e.g. from enteric apicomplexa oocysts
CC   such as Cryptosporidium oocysts and Toxoplasma oocysts.
XX
SQ   Sequence 500 BP; 118 A; 118 C; 131 G; 133 T; 0 other;


  Query Match             47.1%;  Score 21.2;  DB 20;  Length 500;
  Best Local Similarity   76.5%;  Pred. No. 14;
  Matches   26;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      6 aagcttgcggccgatcaggccgaatatgcggccg 39
            ||  || | ||||||| |||||||  |||||| |
Db    473 AATGTTCCAGCCGATCGGGCCGAACTTGCGGCGG 440
```

```
RESULT   4
X21501/c
ID   X21501 standard; DNA; 3481 BP.
XX
AC   X21501;
XX
DT   21-MAY-1999  (first entry)
XX
DE   DNA sequence of sorbitol dehydrogenase (SLDH) and ORF2 genes.
XX
KW   Sorbitol dehydrogenase; SLDH; open reading frame; ORF2; L-sorbose;
KW   D-sorbitol; vitamin C; ds.
XX
OS   Gluconobacter suboxydans.
XX
FH   Key             Location/Qualifiers
FT   RBS             177..182
FT                   /*tag=  a
FT                   /note= "putative Shine-Dalgarno (SD) sequence for
FT                         ORF2 gene"
FT   CDS             192..572
FT                   /*tag=  b
FT                   /product= "ORF2 gene product"
FT                   /note= "corresponding protein sequence shown in W95020"
FT   RBS             558..564
FT                   /*tag=  c
FT                   /note= "putative Shine-Dalgarno (SD) sequence for
FT                         SLDH gene"
FT   CDS             572..2794
FT                   /*tag=  d
FT                   /product= "SLDH gene product"
FT                   /note= "corresponding protein sequence shown in W95019"
FT   sig_peptide     572..643
FT                   /*tag=  e
FT                   /note= "signal sequence for SLDH"
FT   mat_peptide     644..2791
FT                   /*tag=  f
FT                   /note= "mature SLDH protein sequence"
FT   repeat_region   684..704
FT                   /*tag=  g
FT                   /rpt_type= inverted
FT                   /note= "inverted repeat sequence IR1 as possible
FT                          transcription terminator for ORF2 gene as
FT                          indicated in the specification"
FT   repeat_unit     684..693
FT                   /*tag=  h
FT   repeat_region   2803..2892
FT                   /*tag=  i
FT                   /rpt_type= inverted
FT                   /note= "inverted repeat sequence IR2 as possible
FT                          transcription terminator for SLDH gene as
FT                          indicated in the specification"
FT   repeat_unit     2803..2833
FT                   /*tag=  j
XX
PN   EP897984-A2.
XX
PD   24-FEB-1999.
XX
PF   13-AUG-1998;   98EP-0115231.
XX
PR   21-AUG-1997;   97EP-0114432.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Hosino T,  Miyazaki T,  Ojima S,  Shinjoh M,  Tomiyama N;
XX
DR   WPI; 1999-134646/12.
DR   P-PSDB; W95019, W95020.
XX
PT   New D-sorbitol dehydrogenase gene and recombinant protein - useful
PT   for production of L-sorbose, an intermediate in vitamin C production
XX
PS   Claim 4; Fig 3A-D; 39pp; English.
XX
CC   This represents the DNA sequence of the SLDH gene encoding a protein with
CC   sorbitol dehydrogenase (SLDH) activity. The DNA also encodes an open
CC   reading frame (ORF2) product upstream of the SLDH open reading frame,
CC   needed for SLDH activity in vivo. Host cells transformed by a vector
CC   comprising the SLDH DNA sequence are used for the recombinant expression
CC   of the sorbitol dehydrogenase. Recombinant D-sorbitol dehydrogenase is
CC   useful for producing L-sorbose from D-sorbitol. L-sorbose is an important
CC   intermediate in vitamin C production.
XX
SQ   Sequence 3481 BP; 674 A; 1018 C; 1001 G; 788 T; 0 other;


  Query Match             47.1%;  Score 21.2;  DB 20;  Length 3481;
  Best Local Similarity   76.5%;  Pred. No. 18;
  Matches   26;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       4 ccaagcttgcggccgatcaggccgaatatgcggc 37
           |||||| ||||||| || ||   ||||| || ||
Db     388 CCAAGCGTGCGGCCCATGAGCATGAATACGCCGC 355


RESULT   5
V15518
ID   V15518 standard; DNA; 1503 BP.
XX
AC   V15518;
XX
DT   22-MAY-1998  (first entry)
XX
DE   Hamster oral papilloma virus L1 DNA.
XX
KW   L1 DNA; L1 protein; diagnosis; papilloma virus infectious disease;
KW   cancer; screening; antitumour agent; antisense treatment; ss.
XX
OS   Hamster oral papilloma virus.
XX
PN   JP10042875-A.
XX
PD   17-FEB-1998.
XX
PF   06-AUG-1996;   96JP-0207143.
XX
PR   06-AUG-1996;   96JP-0207143.
XX
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 1998-186869/17.
DR   P-PSDB; W47224.
XX
PT   DNA encoding hamster oral papilloma virus protein - useful for, e.g.
PT   diagnosing cancer related papilloma virus infections
XX
PS   Claim 10; Pages 13-14; 17pp; Japanese.
XX
CC   The present sequence is the hamster oral papilloma virus L1 DNA.
CC   The L1 protein and DNA can be used for the diagnosis of papilloma
CC   virus infectious diseases related to cancer, screening for
CC   antitumour agents and antisense treatment.
XX
SQ   Sequence 1503 BP; 434 A; 346 C; 337 G; 386 T; 0 other;


  Query Match             46.2%;  Score 20.8;  DB 19;  Length 1503;
  Best Local Similarity   70.0%;  Pred. No. 23;
  Matches   28;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       6 aagcttgcggccgatcaggccgaatatgcggccgcattat 45
             || | |||||| |||||||   || | || |  |||| ||
Db     826 aaacctgcggctgatcaggatcaaaaggctgatgcatcat 865
```

```
RESULT   6
V15519
ID   V15519 standard; DNA; 7647 BP.
XX
AC   V15519;
XX
DT   22-MAY-1998  (first entry)
XX
DE   Hamster oral papilloma virus genomic DNA.
XX
KW   Diagnosis; papilloma virus infectious disease;
KW   cancer; screening; antitumour agent; antisense treatment; ss.
XX
OS   Hamster oral papilloma virus.
XX
PN   JP10042875-A.
XX
PD   17-FEB-1998.
XX
PF   06-AUG-1996;   96JP-0207143.
XX
PR   06-AUG-1996;   96JP-0207143.
XX
PA   (TORA ) TORAY IND INC.
XX
DR   WPI; 1998-186869/17.
XX
PT   DNA encoding hamster oral papilloma virus protein - useful for, e.g.
PT   diagnosing cancer related papilloma virus infections
XX
PS   Claim 11; Pages 14-16; 17pp; Japanese.
XX
CC   The present sequence is a hamster oral papilloma virus genomic
CC   DNA. The DNA can be used for the diagnosis of papilloma virus
CC   infectious diseases related to cancer, screening for antitumour
CC   agents and antisense treatment.
XX
SQ   Sequence 7647 BP; 2102 A; 1703 C; 1894 G; 1948 T; 0 other;


  Query Match             46.2%;  Score 20.8;  DB 19;  Length 7647;
  Best Local Similarity   70.0%;  Pred. No. 28;
  Matches   28;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       6 aagcttgcggccgatcaggccgaatatgcggccgcattat 45
           || | |||||| |||||||   || | || |  |||| ||
Db    6474 aaacctgcggctgatcaggatcaaaaggctgatgcatcat 6513


RESULT   7
Z36926
ID   Z36926 standard; DNA; 8907 BP.
XX
AC   Z36926;
XX
DT   13-MAR-2000  (first entry)
XX
DE   Nucleotide sequence of the genome of Stealth virus clone 313.
XX
KW   Prototype Stealth virus clone; atypically structured virus;
KW   vacuolating cytopathic effect; cytopathic virus; virus detection;
KW   malignancy; multiple myeloma; lymphoma; brain tumour; breast cancer;
KW   salivary gland tumour; Alzheimer's disease; Parkinson's disease;
KW   spongiform encephalopathy; multiple sclerosis; schizophrenia;
KW   manic depression; major depression; personality disorder; autism;
KW   Rett's syndrome; attention deficit; oppositional defiance; aggression;
KW   anorexia nervosa; bulimia; multi-system illness; virus infection; ss.
XX
OS   Stealth virus.
XX
PN   WO9960101-A1.
XX
PD   25-NOV-1999.
XX
PF   19-MAY-1999;   99WO-US11185.
XX
PR   19-MAY-1998;   98US-0081708.
XX
PA   (MART/) MARTIN W J.
XX
PI   Martin WJ;
XX
DR   WPI; 2000-072436/06.
XX
PT   Isolated viruses, used to develop products for detection, prevention
PT   and treatment of stealth virus infections  -
XX
PS   Disclosure; Page 34-37; 50pp; English.
XX
CC   Z36925-30 represent the nucleotide sequences of prototype Stealth virus
CC   clones. The sequences illustrate the type of sequences that can be used
CC   to generate probes and to predict antigenic and biologically active
CC   products obtainable from a viral isolate, in the methods of the
CC   invention. The specification describes tissue culture, serological and
CC   molecular based methods to detect atypically structured viruses,
CC   such as Stealth viruses, which are able to induce a vacuolating
CC   cytopathic effect (CPE) in tissue culture, and are distinguishable
CC   from known cytopathic viruses, by appearance progression and/or host
CC   range, or by serological, electron microscopic and/or molecular markers.
CC   The products can be used for detecting viruses in patients with
CC   diseases such as a malignancy, e.g. multiple myeloma, lymphoma, brain
CC   tumours, breast cancer, salivary gland tumours, Alzheimer's disease,
CC   Parkinson's disease, spongiform encephalopathy, multiple sclerosis,
CC   schizophrenia, manic depression, major depression, personality
CC   disorders, autism, Rett's syndrome, attention deficit, oppositional
CC   defiance, aggression, anorexia nervosa, bulimia, a multi-system
CC   illness, an animal illness or an illness in a domestic dog or cat.
CC   The products can be used for detecting Stealth viruses in biological
CC   products such as blood products and foods. The products can also be
CC   used for developing agents for treating or preventing virus infections.
XX
SQ   Sequence 8907 BP; 2257 A; 2447 C; 2165 G; 2023 T; 15 other;


  Query Match             46.2%;  Score 20.8;  DB 21;  Length 8907;
  Best Local Similarity   73.5%;  Pred. No. 28;
  Matches   25;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       5 caagcttgcggccgatcaggccgaatatgcggcc 38
             ||   | | ||||||||||||||||  ||:| |||
Db    6014 caccatcggggccgatcaggccgaccatrcagcc 6047


RESULT   8
X84323
ID   X84323 standard; DNA; 9181 BP.
XX
AC   X84323;
XX
DT   08-SEP-1999  (first entry)
XX
DE   Stealth virus nucleic acid clone, SEQ ID NO: 15.
XX
KW   Stealth virus; detection; diagnosis; infection; ss.
XX
OS   Stealth virus.
XX
FH   Key             Location/Qualifiers
FT   misc_difference 8920
FT                   /*tag=  a
FT                   /note= "this nucleotide is represented as a * in the
FT                          specification, and is included to maintain the
FT                          base numbering given in the specification"
FT   misc_difference 8929
```

```
FT                  /*tag= b
FT                  /note= "this nucleotide is represented as a * in the
FT                         specification, and is included to maintain the
FT                         base numbering given in the specification"
FT   misc_difference 8941
FT                  /*tag= c
FT                  /note= "this nucleotide is represented as a * in the
FT                         specification, and is included to maintain the
FT                         base numbering given in the specification"
FT   misc_feature    8945..8947
FT                  /*tag= d
FT                  /note= "these nucleotides are represented as * in the
FT                         specification, and are included to maintain the
FT                         base numbering given in the specification"
FT   misc_feature    8952..8954
FT                  /*tag= e
FT                  /note= "these nucleotides are represented as * in the
FT                         specification, and are included to maintain the
FT                         base numbering given in the specification"
FT   misc_difference 8979
FT                  /*tag= f
FT                  /note= "this nucleotide is represented as a * in the
FT                         specification, and is included to maintain the
FT                         base numbering given in the specification"
FT   misc_difference 9097
FT                  /*tag= g
FT                  /note= "this nucleotide is represented as a * in the
FT                         specification, and is included to maintain the
FT                         base numbering given in the specification"
XX
PN   WO9934019-A1.
XX
PD   08-JUL-1999.
XX
PF   30-DEC-1998;   98WO-US27744.
XX
PR   30-DEC-1997;   97US-0001184.
XX
PA   (MART/) MARTIN W J.
XX
PI   Martin WJ;
XX
DR   WPI; 1999-405521/34.
XX
PT   Novel strains of stealth virus
XX
PS   Claim 19; Page 49-53; 95pp; English.
XX
CC   This sequence represents a Stealth virus nucleic acid clone. The
CC   invention relates to a method of detecting and characterising a stealth
CC   virus by reacting a sample suspected of containing a stealth virus with a
CC   probe from a known stealth virus and sequencing the resultant isolated
CC   nucleotide. The method comprises the steps of: (a) isolating DNA or RNA
CC   from a sample suspected of containing a stealth virus, e.g. a culture of
CC   cells showing a viral cytopathic effect; (b) testing the reactivity of
CC   the isolated DNA or RNA with a molecular probe that contains at least 18
CC   or more contiguous nucleotides identical to sequence previously
CC   identified from a stealth virus; and, optionally (c) sequencing the
CC   isolated DNA or RNA molecules that react with the probe. The method is
CC   used to detect stealth virus in a biological product, food or in the
CC   environment. The method is also used to evaluate agents for their
CC   inhibitory or stimulatory effects on stealth virus replication and to
CC   determine capacity of the virus to recombine with and potentially alter
CC   the nucleic acid sequences of a cell or bacterium.
XX
SQ   Sequence 9181 BP; 2287 A; 2465 C; 2173 G; 2050 T; 206 other;


  Query Match             46.2%;  Score 20.8;  DB 20;  Length 9181;
  Best Local Similarity   73.5%;  Pred. No. 28;
  Matches   25;  Conservative    1;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      5 caagcttgcggccgatcaggccgaatatgcggcc 38
          ||||  | | |||||||||||||||  ||:| |||
Db   6020 caccatcggggccgatcaggccgaccatrcagcc 6053


RESULT   9
Q70114/c
ID   Q70114 standard; DNA; 41 BP.
XX
AC   Q70114;
XX
DT   10-MAR-1995  (first entry)
XX
DE   Lambda INNER primer.
XX
KW   Enterokinase; EK; heavy chain; light chain; catalytic domain;
KW   digestive disorder; cleavage; fusion protein; trypsinogen;
KW   trypsin; enzyme; PACE gene; ss.
XX
OS   Synthetic.
XX
PN   WO9416083-A.
XX
PD   21-JUL-1994.
XX
PF   13-JAN-1994;   94WO-US00616.
XX
PR   15-JAN-1993;   93US-0005944.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Lavallie ER;
XX
DR   WPI; 1994-249229/30.
XX
PT   New nucleic acid encoding enterokinase activity - and related
PT   vectors, host cells, expression products and antibodies are
PT   useful in treating digestive disorders and for cleaving fusion
PT   proteins
XX
PS   Example 6; Page 19; 50pp; English.
XX
CC   Nested oligonucleotide primers were synthesised which were
CC   complementary to the lambda DNA sequence adjacent to the cloning site
CC   for the cDNA insertions (see Q70113-14).  In addition, primers were
CC   designed which were complementary to the plus strand of the most 5'
CC   region of the EK coding sequence (see Q70115-16). Subcloning and
CC   sequencing provided a 116 amino acid EK heavy chain.   This protein
CC   contained regions corresp. to tryptic peptides (see R57295-9).
XX
SQ   Sequence 41 BP; 12 A; 10 C; 11 G; 8 T; 0 other;


  Query Match             45.3%;  Score 20.4;  DB 15;  Length 41;
  Best Local Similarity   71.1%;  Pred. No. 22;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy      8 gcttgcggccgatcaggccgaatatgcggccgcattat 45
            |||||      | |||||  |||  ||||||||||| |
Db     41 GCTTGGACTTAACCAGGCTGAACTTGCGGCCGCATTCT 4


RESULT  10
Z29251
ID   Z29251 standard; DNA; 939 BP.
XX
AC   Z29251;
XX
DT   28-FEB-2000  (first entry)
XX
DE   Rifin 3193 gene.
XX
KW   Rifin 3193 gene; P.falciparum chromosome 2; Rifin protein;
KW   mammalian expression plasmid VR1050; DNA vaccine; immunisation;
```

KW stage specific protein expression; antimicrobial vaccine;
KW antimicrobial drug; IAF; immunofluorescent antibody testing; ds.
XX
OS Plasmodium falciparum.
XX
PN WO9955381-A1.
XX
PD 04-NOV-1999.
XX
PF 26-APR-1999; 99WO-US09047.
XX
PR 24-APR-1998; 98US-0082947.
PR 23-APR-1999; 99US-0082947.
XX
PA (USNA) US SEC OF NAVY.
XX
PI Hoffman SL, Carucci DJ;
XX
DR WPI; 2000-086380/07.
XX
PT Use of microbial, animal and/or human genomic data for identification
PT of vaccine, drug or diagnostic products -
XX
PS Claim 13; Page 16; 23pp; English.
XX
CC The present sequence is the Rifin 3193 gene of chromosome 2. Rifin genes
CC were used to characterise the protein expression from various life cycle
CC stages of P.falciparum. Oligonucleotide primers were used to amplify each
CC selected ORF from chromosome 2 using genomic DNA as template. The
CC purified amplified products were cloned into the mammalian expression
CC plasmid VR1050 to produce DNA vaccines for immunisation. Blood and sera
CC obtained from groups of mice immunised with doses of vaccine was used to
CC identify stage specific expression of Rifin protein by immunofluorescent
CC antibody testing. The method is useful for determination of subcellular
CC localisation of proteins and for the development of antimicrobial
CC vaccines and drugs.
XX
SQ Sequence 939 BP; 365 A; 133 C; 172 G; 269 T; 0 other;

```
  Query Match             45.3%;  Score 20.4;  DB 21;  Length 939;
  Best Local Similarity   71.1%;  Pred. No. 31;
  Matches   27;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy       2 atccaagcttgcggccgatcaggccgaatatgcggccg 39
           ||||||||  || |||| |   || || ||||| || |
Db     702 atccaagcgagctgccggtattgcagattatgctgcag 739
```

RESULT 11
N81768/c
ID N81768 standard; DNA; 4260 BP.
XX
AC N81768;
XX
DT 29-DEC-1990 (first entry)
XX
DE Sequence encoding Mycobacterium tuberculosis 540 and 517 AA residue
DE proteins.
XX
KW Diagnosis; assay; M.bovis; vaccine; ds.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT CDS 252..1874
FT /*tag= a
FT /label=540 AA protein
FT /note="P81351"
FT CDS complement (3948..2395)
FT /*tag= b
FT /label=517 AA protein
FT /note="P81868"
XX
PN WO8806591-A.
XX
PD 07-SEP-1988.
XX
PF 25-FEB-1988; 88WO-US00598.
XX
PR 24-FEB-1988; 88US-0159667.
PR 06-FEB-1987; 87US-0019529.
XX
PA (SCRI-) SCRIPPS CLINIC & RE.
XX
PI Shinnick T, Houghten R;
XX
DR WPI; 1988-271136/38.
DR P-PSDB; P81351, P81868.
XX
PT Recombinant mycobacterial peptide(s) -
PT used in assays for diagnosis of infection, for producing
PT vaccines and for producing antibodies
XX
PS Disclosure; Fig 2a-2d; 116pp; English.
XX
CC An isolated DNA molecule that consists essentially of the nucleotide
CC sequence that corresponds to the sequence represented by position 3950 to
CC about 2390 and from position 3948 through position 2398 of N81768 is
CC claimed. Also claimed is a peptide sequence that consists of a 5-40 AA
CC residue sequence that corresponds to a sequence of the 540 AA residue
CC protein (P81351) or the 517 AA residue protein (P81868) coded for by the
CC DNA sequence. The proteins can be used for determining previous
CC immunological exposure of a mammal to M.tuberculosis or M.bovis and
CC for producing a vaccine.
XX
SQ Sequence 4260 BP; 733 A; 1332 C; 1481 G; 714 T; 0 other;

```
  Query Match             45.3%;  Score 20.4;  DB 9;  Length 4260;
  Best Local Similarity   80.0%;  Pred. No. 37;
  Matches   24;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      12 gcggccgatcaggccgaatatgcggccgca 41
           ||||| |  ||||| ||| | |||||||||
Db    3510 GCGGCAGCGCAGGCGGAACAAGCGGCCGCA 3481
```

RESULT 12
N80222/c
ID N80222 standard; DNA; 4380 BP.
XX
AC N80222;
XX
DT 19-MAR-1991 (first entry)
XX
DE Sequence of Mycobacterium tuberculosis DNA contg. gene encoding 65
DE protein.
XX
KW Antigen; vaccine; ds.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT CDS 192..1874
FT /*tag= a
FT CDS complement (2398..4101)
FT /*tag= b
XX
PN WO8805823-A.
XX
PD 11-AUG-1988.
XX
PF 01-FEB-1988; 88WO-US00281.
XX
PR 02-FEB-1987; 87US-0010007.

XX
PA (WHIT-) WHITEHEAD INST BIOM.
XX
PI Husson RN, Young RA, Shinnick TM;
XX
DR WPI; 1988-235175/33.
DR P-PSDB; P80215, P80216
XX
PT Genes encoding Mycobacterium tuberculosis protein antigens -
PT useful for developing reagents for diagnosis, prevention and
PT treatment of tuberculosis
XX
PS Claim 12; Fig 8; 82pp; English.
XX
CC The gene was isolated by probing a lambda gt11 expression library of
CC M.tuberculosis DNA with monoclonal antibodies directed against
CC M.tuberculosis-specific antigens. The 19kD, 71kD and the 65kD proteins
CC and genes are claimed, and so is a vaccine comprising DNA encoding
CC M.tuberculosis protein in a recombinant vaccine vector. P80216 is
CC encoded on the complementary strand.
XX
SQ Sequence 4380 BP; 757 A; 1373 C; 1512 G; 738 T; 0 other;

Query Match 45.3%; Score 20.4; DB 9; Length 4380;
Best Local Similarity 80.0%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 12 gcggccgatcaggccgaatatgcggccgca 41
|||||| | ||||| ||| | |||||||||
Db 3630 GCGGCAGCGCAGGCGGAACAAGCGGCCGCA 3601

RESULT 13
V05708/c
ID V05708 standard; DNA; 4380 BP.
XX
AC V05708;
XX
DT 22-JUN-1998 (first entry)
XX
DE Mycobacterium tuberculosis 65 kDa heat shock protein gene.
XX
KW Heat shock protein; Mt Hsp65; autoimmune disease; immunotherapy;
KW gene therapy; rheumatoid arthritis; multiple sclerosis; ds.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT CDS 252..1874
FT /*tag= a
FT /product= 65 kDa heat shock protein
XX
PN WO9746253-A2.
XX
PD 11-DEC-1997.
XX
PF 03-JUN-1997; 97WO-US09427.
XX
PR 03-JUN-1997; 97US-0019100.
PR 03-JUN-1996; 96US-0019100.
XX
PA (AURA-) AURAGEN INC.
XX
PI Haynes JR, Prayaga SK, Ramshaw IA;
XX
DR WPI; 1998-041892/04.
DR P-PSDB; W44702.
XX
PT Treatment of autoimmune diseases - by administering
PT autoantigen-coated particles or autoantigen-encoding nucleic acid
PT construct
XX
PS Example 2; Page 55-59; 72pp; English.
XX
CC This DNA sequence encodes the 65 kDa heat shock protein (see
CC W44702), designated Mt Hsp65, of Mycobacterium tuberculosis. This
CC protein cross-reacts with a component of articular cartilage, human
CC Hsp60, that is up-regulated in the joints of arthritic patients. A
CC claimed method for treating or preventing an autoimmune disease in
CC a mammal comprises: (a) providing a particle coated with an antigen
CC against which an immune response is mounted in the autoimmune
CC disease; (b) delivering the particle into the recipient cell of the
CC mammal; and (c) repeating step (b) until either a reduction in a
CC cytotoxic immune response or a desensitizing immune response is
CC induced in the mammal. Alternatively, step (a) comprises providing
CC a nucleic acid construct comprising a coding sequence for the
CC antigen, operably linked to control elements such that the coding
CC sequence can be transcribed and translated in a recipient cell, and
CC delivering the construct to the recipient cell using a gene gun.
CC The antigen of step (a) is selected from collagen, Mt Hsp65,
CC myelin basic protein, myelin oligodendrocyte glycoprotein,
CC proteolipid protein, and epitopes thereof. These antigens mitigate
CC cytotoxic responses and elicit antigen desensitisation. The method
CC is used especially for treating rheumatoid arthritis or multiple
CC sclerosis. It represents a novel use for the known Mt Hsp65 gene.
XX
SQ Sequence 4380 BP; 757 A; 1371 C; 1514 G; 738 T; 0 other;

Query Match 45.3%; Score 20.4; DB 19; Length 4380;
Best Local Similarity 80.0%; Pred. No. 37;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 12 gcggccgatcaggccgaatatgcggccgca 41
|||||| | ||||| ||| | |||||||||
Db 3630 GCGGCAGCGCAGGCGGAACAAGCGGCCGCA 3601

RESULT 14
Z46842/c
ID Z46842 standard; cDNA; 1818 BP.
XX
AC Z46842;
XX
DT 10-APR-2000 (first entry)
XX
DE Semaphorin K1 gene related sequence.
XX
KW Semaphorin K1; cellular physiology; neurite outgrowth; neuron; human;
KW immunogen; pharmaceutical; ss.
XX
OS Unidentified.
XX
PN JP11341988-A.
XX
PD 14-DEC-1999.
XX
PF 11-MAR-1999; 99JP-0065672.
XX
PR 11-MAR-1998; 98US-0041236.
XX
PA (EXEL-) EXELIXIS PHARM INC.
XX
DR WPI; 2000-109378/10.
DR P-PSDB; Y56855.
XX
PT New semaphorin polypeptides, useful cell physiology modulators and
PT immunogens -
XX
PS Disclosure; Page 17-20; 57pp; Japanese.
XX
CC The invention provided isolated human semaphorin K1 polypeptides. The
CC polypeptides, or nucleic acids encoding them, can be used to modulate
CC cellular physiology by modulating semaphorin K1 activity, e.g. semaphorin
CC K1 polypeptide fragments or antisense nucleic acids can be used to

CC enhance neurite outgrowth from damaged neurons. The polypeptides can also
CC be used as immunogens, reagents for isolating other semaphorins, or as
CC reagents for screening chemical libraries for lead pharmaceutical agents.
CC The nucleic acids can also be used as probes and primers for diagnostic
CC purposes.
XX
SQ Sequence 1818 BP; 543 A; 409 C; 359 G; 507 T; 0 other;

Query Match 44.9%; Score 20.2; DB 21; Length 1818;
Best Local Similarity 68.3%; Pred. No. 40;
Matches 28; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 5 caagcttgcggccgatcaggccgaatatgcggccgcattat 45
|||||| || ||| ||| | | | |||| || | ||||
Db 353 CAAGCTAGCAGCCCATCTGTCTGGTTATGAAGCAGAGTTAT 313

RESULT 15
X40055
ID X40055 standard; DNA; 2885 BP.
XX
AC X40055;
XX
DT 02-JUL-1999 (first entry)
XX
DE Colon cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 651-652; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
XX
SQ Sequence 2885 BP; 626 A; 901 C; 838 G; 520 T; 0 other; .

Query Match 44.9%; Score 20.2; DB 20; Length 2885;
Best Local Similarity 75.8%; Pred. No. 43;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 8 gcttgcggccgatcaggccgaatatgcggccgc 40
| | | |||||| |||||| || |||||| ||
Db 2715 ggtgggggccgaacaggcccaagctgcggcagc 2747

Search completed: March 4, 2001, 12:07:04
Job time: 5659 sec